8-14-2014

Modeling, Simulation and Fabrication of 100 nm (\textit{Leff}) High Performance CMOS Transistors

Chandan K. Amareshbabu

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation
Modeling, Simulation and Fabrication of 100 nm ($L_{eff}$) High Performance CMOS Transistors

By
Chandan K. Amareshbabu
A Thesis Submitted
In Partial Fulfilment
Of the Requirements for the Degree of
Master of Science
In
Electrical Engineering

Approved by:
Professor: ____________________________
Dr. Lynn F. Fuller (Thesis Advisor)
Professor: ____________________________
Dr. Karl D. Hirschman (Thesis Committee Member)
Professor: ____________________________
Dr. Ivan Puchades (Thesis Committee Member)
Professor: ____________________________
Dr. Sohail A. Dianat (Department Head)

DEPARTMENT OF ELECTRICAL AND MICROELECTRONICS ENGINEERING
KATE GLEASON COLLEGE OF ENGINEERING
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NY
AUGUST 14, 2014
Modeling, Simulation and Fabrication of 100 nm ($L_{eff}$) High Performance CMOS Transistors

By,

Chandan K. Amareshbabu

I, Chandan Amareshbabu, hereby grant permission to the Wallace Memorial Library of Rochester Institute of Technology to reproduce this document in whole or in part, such that any reproduction will not be for commercial use or profit.

______________________________  08/14/2014

Chandan Amareshbabu  Date
I dedicate this thesis to my parents, Amareshbabu and Sreelaxmi who have been a constant source of support to me.
ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Lynn Fuller for the continuous support in the research with his immense knowledge, motivation and enthusiasm. He has had a positive influence on me throughout this work. I have gained a lot of knowledge from your Microelectronics Manufacturing and MEMS courses.

I would also like to thank Dr. Hirschman and Dr. Ivan for their continued support by providing advice through this work. Dr. Hirschman’s device modeling course set a baseline to my research and Dr. Ivan’s MEMS fabrication course provided fabrication experience in SMFL.

I would also want to thank Dr. Robert Pearson, Dr. Kurinec, Dr. Michael Jackson and Dr. James Moon. I gained a lot of knowledge through your coursework and learnt a lot of new concepts over two years at RIT.

I would like to thank staff of Semiconductor and Microsystems Fabrication Laboratory specially Patricia Meller and Sean O’Brien for constant help while working in laboratory. I would like to thank Scott Blondell, John Nash, Bruce Tolleson, Rich Battaglia and Dave Yackoff for training and certifying on Varian 350D Implanter; ASML stepper, P5000, CVC evaporator, LAM 4600; LPCVD, DryTek Quad, SSI track; LAM 490, Implanter; wet chemical benches, RTA tool and Bruce furnace respectively.

Finally, I would like to specially thank graduate student Samarth Parikh who worked with me throughout this work in the Fab and outside Fab. Abhinav Gaur, who helped us in using LEO SEM to obtain beautiful pictures of Polysilicon and Nitride.
ABSTRACT

A 100 nm CMOS process is modeled and simulated using advanced ion implantation, diffusion models, recombination and mobility models; in ATHENA and ATLAS respectively. This process is designed for a mask length of 150 nm and effective gate length of =<100 nm.

The drive current on modeled and simulated device for NMOS and PMOS are 300 μA/μm and 63 μA/μm at 1.2 V drain and gate bias respectively. The threshold voltage for the modeled NMOS and PMOS is 0.479 V and -0.58 V with DIBL less than 12 mV/V on both the devices.

Unit process improvements includes retrograde well doping to reduce short channel effects, double exposure to obtain minimum feature size, anisotropic polysilicon and nitride spacers etch, high dose source drain extension implants to minimize the contribution of parasitic resistance to ON state performance of the device, high dose source drain implants followed by SALICIDE process to provide ohmic contacts. Fabricated devices failed because of excessive gate leakage.
## TABLE OF CONTENTS

### CHAPTER 1 – INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

### CHAPTER 2 – MOSFET BACKGROUND

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 MOS STRUCTURES</td>
<td>3</td>
</tr>
<tr>
<td>2.2 THRESHOLD VOLTAGE</td>
<td>8</td>
</tr>
<tr>
<td>2.3 SHORT CHANNEL EFFECTS</td>
<td>9</td>
</tr>
<tr>
<td>2.3.1 CHANNEL LENGTH MODULATION</td>
<td>9</td>
</tr>
<tr>
<td>2.3.2 PUNCH THROUGH</td>
<td>9</td>
</tr>
<tr>
<td>2.3.3 DRAIN INDUCED BARRIER LOWERING</td>
<td>10</td>
</tr>
<tr>
<td>2.3.4 THRESHOLD VOLTAGE ROLL OFF</td>
<td>11</td>
</tr>
<tr>
<td>2.4 SCALING PARAMETERS OF NTRS</td>
<td>11</td>
</tr>
<tr>
<td>2.4.1 PHYSICAL PARAMETERS</td>
<td>11</td>
</tr>
<tr>
<td>2.4.2 ELECTRICAL PARAMETERS</td>
<td>13</td>
</tr>
<tr>
<td>2.4.3 DESIGN PARAMETERS</td>
<td>14</td>
</tr>
<tr>
<td>2.4.3.1 FIRST ORDER CALCULATION DESIGN</td>
<td>14</td>
</tr>
<tr>
<td>2.4.3.1.1 BUILT IN POTENTIAL</td>
<td>14</td>
</tr>
<tr>
<td>2.4.3.1.2 WIDTH OF SPACE CHARGE REGION</td>
<td>15</td>
</tr>
<tr>
<td>2.4.3.1.3 EFFECTIVE GATE LENGTH</td>
<td>17</td>
</tr>
<tr>
<td>2.4.3.2 ATHENA AND ATLAS</td>
<td>18</td>
</tr>
</tbody>
</table>

### CHAPTER 3 - DEVICE MODELING AND SIMULATION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 ATHENA</td>
<td>19</td>
</tr>
<tr>
<td>3.2 GRID STRUCTURE</td>
<td>19</td>
</tr>
<tr>
<td>3.3 INITIAL SUBSTRATE</td>
<td>20</td>
</tr>
<tr>
<td>3.4 ION IMPLANTATION MODELS</td>
<td>21</td>
</tr>
<tr>
<td>3.4.1 GAUSSIAN MODEL</td>
<td>21</td>
</tr>
<tr>
<td>3.4.2 PEARSON MODEL</td>
<td>22</td>
</tr>
<tr>
<td>3.4.3 DUAL PEARSON MODEL</td>
<td>22</td>
</tr>
<tr>
<td>3.4.4 SIMS VERIFIED DUAL PEARSON MODEL</td>
<td>23</td>
</tr>
<tr>
<td>3.4.5 MONTE CARLO IMPLANT MODEL</td>
<td>24</td>
</tr>
<tr>
<td>3.4.5.1 NUCLEAR STOPPING</td>
<td>24</td>
</tr>
<tr>
<td>3.4.5.2 INTERATOMIC POTENTIAL</td>
<td>25</td>
</tr>
<tr>
<td>3.4.5.3 ELECTRONIC STOPPING</td>
<td>25</td>
</tr>
</tbody>
</table>
3.4.5.4 DAMAGE ACCUMULATION MODEL ............................................................... 26
3.5 DEPOSITION MODEL ........................................................................................... 27
3.6 ETCHING MODEL ............................................................................................... 28
3.7 DIFFUSION MODELS .......................................................................................... 29
  3.7.1 FERMI DIFFUSION MODEL ............................................................................ 29
  3.7.2 IMPURITY SEGREGATION MODEL .............................................................. 30
  3.7.3 TWO DIMENSIONAL MODEL ..................................................................... 31
  3.7.4 FULLY COUPLED MODEL ......................................................................... 32
  3.7.5 RAPID THERMAL ANNEALING ................................................................. 33
3.8 ATLAS .................................................................................................................. 36
  3.8.1 LOMABARDI CVT MODEL ......................................................................... 36
  3.8.2 SHOCKLEY READ HALL MODEL (SRH) ..................................................... 36

CHAPTER 4 - UNIT PROCESS IMPROVEMENTS
4.1 WELL, RETROGRADE AND SOURCE DRAIN EXTENSION IMPLANT .... 37
4.2 GATE OXIDE GROWTH .................................................................................... 37
4.3 DOUBLE EXPOSURE .......................................................................................... 38
4.4 POLYSILICON AND NITRIDE ETCH (RIE) ....................................................... 38
4.5 LOW RESISTANCE SOURCE DRAIN CONTACTS ........................................... 41
4.6 SPIKE ANNEAL ................................................................................................ 41
4.7 TITANIUM SILICIDE ....................................................................................... 42

CHAPTER 5 - FABRICATION OF 100 nm (Leff) CMOS DEVICES
5.1 INTRODUCTION ................................................................................................. 44
5.2 STARTING WAFER ............................................................................................. 45
5.3 PAD OXIDE GROWTH ...................................................................................... 45
5.4 NITRIDE DEPOSITION ...................................................................................... 46
5.5 FIRST LEVEL PHOTO (RECESSES OXIDE) .................................................... 47
5.6 NITRIDE ETCH ................................................................................................ 47
5.7 PHOTORESIST STRIP ....................................................................................... 49
5.8 RCA CLEAN ..................................................................................................... 49
5.9 FIRST OXIDE GROWTH (RECESSED OXIDE) ................................................ 50
5.10 OXIDE ETCH (WET ETCH) ................................................................. 51
5.11 SECOND OXIDE GROWTH (RECESSED OXIDE) ............................... 52
5.12 NITRIDE ETCH (WET ETCH) ....................................................... 52
5.13 SECOND LEVEL PHOTO (NWELL) ............................................... 53
5.14 NWELL IMPLANT ........................................................................ 54
5.15 PHOTORESIST STRIP ................................................................... 54
5.16 THIRD LEVEL PHOTO (PWELL) .................................................. 55
5.17 P-WELL IMPLANT ........................................................................ 55
5.18 PHOTORESIST STRIP ................................................................... 56
5.19 WELL DRIVE-IN ........................................................................... 56
5.20 FOURTH LEVEL PHOTO (N-RETRO) ............................................ 57
5.21 N-RETROGRADE WELL IMPLANT ............................................... 57
5.22 PHOTORESIST STRIP ................................................................... 58
5.23 FIFTH LEVEL PHOTO (P-RETRO) ............................................... 58
5.24 P-RETROGRADE WELL IMPLANT ............................................... 59
5.25 PHOTORESIST STRIP ................................................................... 59
5.26 PAD OXIDE ETCH ........................................................................ 59
5.27 RCA CLEAN ................................................................................. 60
5.28 PRE-OXIDE ETCH TO GATE OXIDE GROWTH .............................. 60
5.29 GATE OXIDE GROWTH ............................................................... 60
5.30 LPCVD POLYSILICON DEPOSITION ............................................. 62
5.31 SIXTH LEVEL PHOTO (POLY)–DOUBLE EXPOSURE .................... 63
5.32 POLYSILICON ETCH .................................................................... 63
5.33 PHOTORESIST STRIP ................................................................... 64
5.34 RCA CLEAN ................................................................................. 64
5.35 POLYSILICON RE-OXIDATION ...................................................... 64
5.36 SEVENTH LEVEL PHOTO (P S/D EXTENSION) ............................ 66
5.37 P+ S/D EXTENSION IMPLANT ...................................................... 66
5.38 PHOTORESIST STRIP ................................................................... 67
5.39 EIGHTH LEVEL PHOTO (N S/D EXTENSION) ................................ 67
5.40 N+ S/D EXTENSION IMPLANT ...................................................... 67
5.41 PHOTORESIST STRIP ................................................................... 68
5.42 RCA CLEAN ................................................................................. 68
6.7 TERADA MUDA ANALYSIS ................................................................................. 93
6.8 SHEET RESISTANCE COMPARISION (FABRICATED VS SIMULATED) ........... 95
6.9 FAILURE OF FABRICATED DEVICES ............................................................... 95

CHAPTER 7 - CONCLUSION AND FUTURE WORK .............................................. 97

APPENDIX A

ATHENA NMOS PROGRAM ................................................................................ 99
NMOS ATLAS IV CURVES PROGRAM (LINEAR) ................................................ 102
NMOS DIBL PROGRAM ....................................................................................... 102
NMOS FAMILY OF CURVES ............................................................................... 103
ATHENA PMOS PROGRAM ................................................................................ 104
PMOS ATLAS IV CURVES PROGRAM (LINEAR) ................................................ 107
PMOS DIBL PROGRAM ....................................................................................... 107
PMOS FAMILY OF CURVES ............................................................................... 108
REFERENCES ........................................................................................................ 109
LIST OF FIGURES

Fig. 1 Industry scaling trends vs RIT scaling trends .......................................................... 1
Fig. 2 Cross section of long channel MOSFET at equilibrium ........................................... 4
Fig. 3 Energy band diagram across source, channel and drain at equilibrium ....................... 4
Fig. 4 Cross section of long channel MOSFET at inversion region of operation ............... 6
Fig. 5 Energy band diagram of inversion region of operation ............................................ 6
Fig. 6 Family of curves showing weak, moderate and strong inversion region of operation .......................................................................................................................... 7
Fig. 7 Family of curves due to CLM .................................................................................... 9
Fig. 8 Family of curves due to punchthrough ...................................................................... 9
Fig. 9 Ideal subthreshold curve and gate leakage curve ...................................................... 10
Fig. 10 Barrier lowering due to DIBL on long channel and short channel ...................... 10
Fig. 11 (V_T roll off) Gate length vs Threshold voltage ...................................................... 11
Fig. 12 (Reverse V_T roll off) gate length vs Threshold voltage ...................................... 11
Fig. 13 Quasi Neutral region and depletion region for drain, channel and source regions for PMOS and NMOS .................................................................................. 16
Fig. 14 Grid structure of silicon on which device is modeled ........................................... 19
Fig. 15 Starting P-Type silicon substrate with <100> orientation .................................... 20
Fig. 16 Conformal deposition of gate oxide using deposition model ................................ 27
Fig. 17 Geometrical etch of polysilicon in ATHENA ...................................................... 28
Fig. 18 Furnace anneal comparision between fermi, two dimensional and fully coupled models .............................................................................................................. 34
Fig. 19 RTA diffusion comparision between fermi, two dimensional and fully coupled models .................................................................................................................. 35
Fig. 20 SEM image of polysilicon gate ............................................................................. 39
Fig. 21 SEM image of nitride spacer ............................................................................. 40
Fig. 22 Initial P-Type substrate with 500Å pad oxide grown on it .................................... 45
Fig. 23 CVD nitride deposition ..................................................................................... 46
Fig. 24 Plot of light sensitivty vs time in seconds .............................................................. 48
Fig. 25 Nitride plasma etch ............................................................................................. 48
Fig. 26 RCA clean process flow ..................................................................................... 50
Fig. 27 First field oxide growth ....................................................................................... 51
Fig. 28 Wet chemical (BOE) thermal oxide etch with steep undercut ............................. 51
Fig. 29 Second field oxide re-growth ............................................................................. 52
Fig. 30 Cross sections of recessed oxide isolation after nitride etch .................................. 53
Fig. 31 High energy NWELL implant ................................................................................... 54
Fig. 32 High energy PWELL implant .................................................................................. 55
Fig. 33 PWELL after drive-in ............................................................................................... 56
Fig. 34 NWELL after drive in ............................................................................................... 57
Fig. 35 N-Type retrograde implant ...................................................................................... 58
Fig. 36 P-Type retrograde implant ...................................................................................... 59
Fig. 37 Gate oxide growth .................................................................................................... 61
Fig. 38 LPCVD polysilicon deposition .................................................................................. 62
Fig. 39 Polysilicon etch ....................................................................................................... 63
Fig. 40 Polysilicon re-ox and final NMOS retrograde profile .............................................. 65
Fig. 41 Polysilicon re-ox and final PMOS retrograde profile .............................................. 65
Fig. 42 PMOS BF2 source drain extension implant ............................................................... 66
Fig. 43 NMOS, Arsenic source drain extension implant ....................................................... 68
Fig. 44 LPCVD nitridde sidewall spacers ............................................................................. 69
Fig. 45 RIE nitride etch ...................................................................................................... 70
Fig. 46 NMOS (P31) Source Drain implant ......................................................................... 71
Fig. 47 PMOS (BF2) Source Drain implant ......................................................................... 73
Fig. 48 NMOS after spike anneal ......................................................................................... 74
Fig. 49 Vertical cutline of NMOS source drain regions and their extensions with concentrations and junction depth .................................................................................................................. 75
Fig. 50 Horizontal cutline of NMOS in channel region ....................................................... 75
Fig. 51 PMOS after spike anneal ......................................................................................... 76
Fig. 52 Vertical cutline of PMOS source drain regions and their extensions with concentrations and junction depth .................................................................................................................. 76
Fig. 53 Horizontal cutline of PMOS in the channel region .................................................... 77
Fig. 54 Titanium deposition ............................................................................................... 78
Fig. 55 Titanium Salicide C49 Phase .................................................................................. 78
Fig. 56 Titanium Salicide C54 phase .................................................................................. 79
Fig. 57 PECVD deposition ................................................................................................. 80
Fig. 58 RIE contact cut etch .............................................................................................. 81
Fig. 59 Aluminum sputter deposition .................................................................................. 83
Fig. 60 Aluminum etch ...................................................................................................... 84
Fig. 61 NMOS linear IV curve ............................................................................................ 87
Fig. 62 PMOS linear IV curve .............................................................88
Fig. 63 NMOS logarithmic DIBL IV curve ..................................................89
Fig. 64 PMOS logarithmic DIBL IV curve .....................................................90
Fig. 65 NMOS family of curves .................................................................91
Fig. 66 PMOS family of curves ................................................................92
Fig. 67 PMOS and NMOS Vt Roll Off ..........................................................93
Fig. 68 NMOS Rm vs Gate Length ..............................................................94
Fig. 69 PMOS Rm vs Gate Length ..............................................................94
Fig. 70 IV curve of failed NMOS fabricated device ....................................95
Fig. 71 IV curve of failed PMOS fabricated device ....................................96

LIST OF TABLES

Table. 1 Comparison of NTRS physical parameters with simulated device at RIT ....12
Table. 2 Comparison of NTRS electrical parameters to simulated device at RIT .....13
Table. 3 Accuracy of SVDP implant model ranges........................................23
Table. 4 Duration of TED at various temperature ranges..............................33
Table. 5 Aluminum etch recipe (LAM4600) ..................................................84
Table. 6 Sheet resistance comparison ..........................................................95
LIST OF ABBREVIATIONS

MOSFET - Metal Oxide Semiconductor Field Effect Transistor
CMOS - Complementary Metal Oxide Semiconductor
NMOS - N-Channel Metal Oxide Semiconductor
PMOS - P-Channel Metal Oxide Semiconductor
CLM - Channel Length Modulation
DIBL - Drain Induced Barrier Lowering
VT - Threshold Voltage
SS - Sub-threshold Slope
NTRS - National Technology Roadmap for Semiconductors
SVDP - SIMS Verified Dual Pearson
BCA - Binary Collision Approximation
RTA - Rapid Thermal Annealing
SRH - Shockley Read Hall
RIE - Reactive Ion Etching
LDD - Low Doped Drain
TEOS - Tetra Ethyl Ortho Silicate
TED - Transient Enhanced Diffusion
UV - Ultra Violet
HMDS - Hexamethyldisilazane
RCA - Radio Corporation of America
BOE - Buffered Oxide Etch
HF - Hydro Fluoric acid
LPCVD - Low Pressure Chemical Vapor Deposition
PECVD - Plasma Enhanced Chemical Vapor Deposition
CMP - Chemical Mechanical Planarization
CHAPTER 1

INTRODUCTION

In 1965, Intel co-founder Gordon Moore predicted that density of transistors on an integrated circuit would double every 18-24 months [1]. His statement has been the driving force for rapid scaling of transistor gate lengths until today’s generation. The transistors with 22 nm gate lengths are in large volume production today (2014) compared to 10 µm in 1971.

For over thirty years, Microelectronics department of RIT has continued to maintain semiconductor industry trend of manufacturing smaller, faster and efficient devices.

The goal of this work is to use the available resources in SMFL to its limits and develop a baseline process to fabricate 100 nm $L_{\text{eff}}$ CMOS transistors at RIT.

A MOSFET is a device which is the core of the integrated circuit design which in conjunction with other circuit elements is capable of voltage gain and signal power gain. Millions of smaller MOSFETs can be integrated or fabricated in a single chip.
Introduction

and has wide range of applications in digital circuits [1]. The lateral operation of MOSFET allows for the reduction of the size and increase in density of devices per unit area and gain in the performance of the MOSFETs.

As seen in the figure, for over thirty years RIT is pushing its limits of available resources to fabricate smallest device possible. The previous work was from Michael Aquilino in 2006 who fabricated 250 nm devices.

The goal of this work is to fabricate 100 nm or <100 nm effective gate length ($L_{eff}$) devices and I have addressed most of the scaling techniques that is used to design this process which encompass the right use of well and source/drain regions dose, retrograde well implant, thin gate oxide growth, shallow source/drain regions, spike annealing techniques, lithography techniques to achieve this gate length and change in Silicide recipe. The use of recessed oxide for isolation is also discussed instead of STI due to the absence of CMP.

The simulation tool used to design this process is SILVACO-SUPREM. ATHENA is used to model the device and ATLAS is used to simulate the electrical characteristics from the modeled device.
CHAPTER 2

MOSFET BACKGROUND

Metal-Oxide-Semiconductor structure is obtained by growing a thin dielectric material, silicon dioxide; upon the semiconducting surface silicon and by depositing metal over the dielectric where the inversion layer is just below the interface of Metal-SiO$_2$-Si interface. MOS capacitor structures are used to study the dielectric material. This research has been a part of university trend to keep up with the fast growing scaling trend of industry to fabricate smaller devices.

2.1 MOS STRUCTURES

There are three types of MOS structures, two terminal, three terminal and four terminal structures.

Two terminal structure is used to study the dielectric (gate oxide) used in the device. The metal in this two terminal structure can be highly conducting polysilicon which is used to obtain the desired work-function or a metal with fixed work-function and hence the threshold voltage of the device. The threshold voltage also depends on the semiconductor doping.

Two terminal structures are used to study the dielectric characteristics before fabricating 75 step long 4 terminal structure.
MOSFET Background

2.1.1 AT EQUILIBRIUM or ACCUMULATION ($V_{GS} < V_T$)

In this region of operation, when the gate bias is negative or almost zero it attracts and accumulates positive charges to the surface and is called as accumulation. The source and the body terminals are grounded.

![Cross section of long channel MOSFET at equilibrium](image)

Fig. 2: Cross section of long channel MOSFET at equilibrium.

The energy band diagram of this region of operation shows that the barrier height $\Phi_{bi}$ stops the major flow of current from source to drain.

![Energy band diagram across source, channel and drain at equilibrium](image)

Fig. 3: Energy band diagram across source, channel and drain at equilibrium.
MOSFET Background

The barrier height is given by

$$\phi_{bi} = \frac{kT}{q} \ln \left( \frac{N_a N_d}{n_i^2} \right)$$

Where, $N_a$ is the acceptor concentration, $N_d$ is donor concentration, $n_i$ is intrinsic concentration, $k$ is Boltzmann constant, $T$ is temperature in Kelvin and $q$ is charge of an electron.

And capacitance across gate and the channel region is given by

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

Where $C_{ox}$ is oxide capacitance, $t_{ox}$ is oxide thickness and $\varepsilon_{ox}$ is permittivity of oxide.

2.1.2 DEPLETION

When positive voltage is applied, the positive charges accumulated in the channel start repelling away, resulting in the formation of depletion region. The total capacitance across gate and the channel region is now the sum of both oxide capacitance and depletion region capacitance.

Capacitance of depletion region is given by,

$$C_D = \frac{\varepsilon_s}{x_d}$$

Where $C_D$ is depletion region capacitance, $x_d$ is depletion region thickness and $\varepsilon_s$ is permittivity of silicon.

The total capacitance is given by

$$C = \frac{1}{C_D} + \frac{1}{C_{ox}}$$
2.1.3 INVERSION ($V_{GS} > V_T$ and $V_{DS}$ is positive)

When there is increase in positive gate charge on the gate the electrons diffuse into the channel from source and when drain bias is applied the negative carriers overcome the source barrier resulting in current in the device.

![Cross section of long channel MOSFET at inversion region of operation.](image)

Fig. 4: Cross section of long channel MOSFET at inversion region of operation.

The energy band diagram shows the lowering of barrier and the inversion of p-type channel region to n-type.

![Energy band diagram of inversion region of operation.](image)

Fig. 5: Energy band diagram of inversion region of operation.

The current at this region of operation is called drive current which evaluates the on state performance of the device and it is given by
MOSFET Background

**Equation [5]**  \[ I_D = \mu_n C_{ox} \frac{W}{L} \left((V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right) \]

Where, \( I_D \) is drain current, \( \mu_n \) is mobility, \( C_{ox} \) is oxide capacitance, \( W \) is width of transistor, \( L \) is length of the transistor, \( V_{GS}, V_T \) and \( V_{DS} \) is gate, threshold and drain voltage respectively.

![Log (I_D) vs V_GB](image.png)

**Fig. 6**: IV curve showing weak, moderate and strong inversion region of operation [2].

When gate bias is less than threshold voltage or almost equal to threshold voltage; that region is called weak inversion. In this mode inversion charge density is less than the magnitude of semiconductor doping concentration [1]. When gate charge is slightly increased and is equal to threshold voltage it is called moderate inversion. In this mode the inversion charge density is almost equal to magnitude of semiconductor doping concentration. Finally at higher gate bias when inversion charge density is greater than magnitude of semiconductor doping concentration it is called strong inversion.
2.2 THRESHOLD VOLTAGE

Threshold voltage is the minimum voltage required to turn on the transistor. At this voltage the barrier height between source and channel region is lowered and allows the flow of carriers into channel region. Threshold voltage of a device depends on several other parameters like doping concentration of polysilicon (metal) and semiconductor (metal semiconductor work-function $\varphi_{ms}$) and interface trapped charges (oxide trapped charges $Q_f$) corresponding to flat band voltage. Oxide capacitance, sum of voltages across the semiconductor ($2\varphi_f$) and depletion layer charge. The mathematical formula of threshold voltage is given by,

$$\text{Equation [6]} \quad V_T = V_{FB} + 2\varphi_f + \frac{2\varepsilon_s q N_A (2\varphi_f + V_{BS})}{C_{ox}}$$

Where, $V_T$ is threshold voltage, $V_{FB}$ is flat band voltage, $\varphi_f$ is potential at source-channel and channel-drain barrier (bulk potential), $\varepsilon_s$ is the permittivity of silicon and free space, $q$ is charge of an electron and $C_{ox}$ is oxide capacitance.

Flat band voltage $V_{FB}$ is given by,

$$\text{Equation [7]} \quad V_{FB} = \varphi_{ms} - \frac{Q_f}{C_{ox}}$$

Where, $\varphi_{ms}$ is metal semiconductor work-function, $Q_f$ is surface charge, $C_{ox}$ is oxide capacitance.

Metal semiconductor work function is given by,

$$\text{Equation [8]} \quad \varphi_{ms} = \varphi_m - \varphi_s$$

Where, $\varphi_m$ is work-function of metal gate used and $\varphi_s$ is work-function of semiconductor (doping concentration dependent).
2.3 SHORT CHANNEL EFFECTS

2.3.1 CHANNEL LENGTH MODULATION

Channel length modulation is the first short channel effect noticed or studied while fabricating small devices.

As it can be noticed in the figure above in the saturation region, there exists a positive slope in the drain current as the drain voltage is increased, resulting decrease in output resistance. Because the increase in drain bias increases the space charge region on the drain and decreasing the length of the inverted channel.

2.3.2 PUNCH THROUGH

Punch through is a short channel effect where the current path is located in the bulk and gate cannot control the flow of charges.
MOSFET Background

This phenomenon increases the sub-threshold leakage current and also sub-threshold slope. The depletion region in the drain increases with the increase in drain voltage, as the depletion region of source and drain approaches it results in current path in the bulk.

![Image of subthreshold curve and gate leakage curve](image)

**Fig. 8:** Ideal subthreshold curve and gate leakage curve.

The figure above is the sub-threshold curve which shows two curves, one is the ideal curve with no short channel effects and other is the curve with punch-through increasing the leakage current.

### 2.3.3 DRAIN INDUCED BARRIER LOWERING

DIBL is a short channel effect where threshold voltage of a transistor reduces at higher drain bias.

![Image of barrier lowering due to DIBL on long channel and short channel](image)

**Fig. 9:** Barrier lowering due to DIBL on long channel and short channel.

The dotted line in the figure is a curve when gate voltage is increased and lowering the energy barrier in the channel region and electrons enter the channel. The dark
MOSFET Background

curve represents a situation when the gate voltage is more than $V_T$ and drain voltage is increased. This lowers the conduction band in the drain region by $qV_{DS}$. Electrons in the source have enough energy to enter the channel and are accelerated by field towards drain. In shorter channel devices this effect further decreases the threshold voltage of the device due to severe charge sharing effects and allowing it to turn on sooner than designed to.

2.3.4 THRESHOLD VOLTAGE ROLL OFF

![Gate length vs Threshold voltage](image)

Fig. 10: ($V_T$ Roll off) Gate length vs Threshold voltage.

The decrease in threshold voltage with decrease in gate length is a common short channel effect. At a very small gate length there is a sharp dip in the threshold voltage and may roll off to zero at a certain gate length operating like a depletion mode device due to surface punch-through.

2.4 SCALING PARAMETERS OF NTRS

2.4.1 PHYSICAL PARAMETERS

Gate oxide is scaled with scaling gate length to give more control over channel to gate. Thermal oxide is used as a dielectric with nitrogen incorporation. Nitrogen incorporation in gate oxide avoids the diffusion of boron from P+ doped polysilicon to channel during anneal. The problem of increased effective thickness of dielectric
due to poly depletion is mitigated by degenerately doping the poly for NMOS and PMOS.

The length of the side wall spacers is defined by thickness of polysilicon deposited. The purpose of side wall spacers in the process is to form S/D extensions, mask the source drain implants and also avoids the short between gate, source and drain during Salicide formation.

<table>
<thead>
<tr>
<th>MODEL PARAMETERS</th>
<th>NTRS (130nm)</th>
<th>NMOS (125nm)</th>
<th>PMOS (125nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate Oxide (nm)</td>
<td>2-3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Side Wall Spacer (nm)</td>
<td>52-104</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Source/Drain Extensions (nm)</td>
<td>50-100</td>
<td>90-100</td>
<td>90-100</td>
</tr>
<tr>
<td>Source/Drain Extension Con(cm⁻³)</td>
<td>1x10¹⁹</td>
<td>8x10¹⁹</td>
<td>1.5x10¹⁹</td>
</tr>
<tr>
<td>Channel Concentration(cm⁻³)</td>
<td>1-2x10¹⁸</td>
<td>1x10¹⁸</td>
<td>5x10¹⁷</td>
</tr>
<tr>
<td>Retrograde well concentration (cm⁻³)</td>
<td>6-10x10¹⁸</td>
<td>5.5x10¹⁸</td>
<td>5x10¹⁸</td>
</tr>
<tr>
<td>Silicide thickness (nm)</td>
<td>40</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Source drain extensions implants well known as low doped drains are not lower doped in sub 100nm technology. Reason for higher doped source drain extensions is to reduce the parasitic resistance and improve on state performance. Also the junction depth of these extensions is very shallow, avoiding short channel effects. Heavy impurities atoms are used to control the lateral diffusion during anneal (BF₂ for PMOS and As for NMOS).
**MOSFET Background**

The high dose deeper source drain implants are used for good ohmic contact reducing the sheet resistance of highly doped N+ and P+ regions by forming Salicide in these regions. Implant energy of deeper junctions are slightly high.

Channel concentration in the surface of the devices is increased to diminish the impact of lateral diffusion of impurities and it is controlled by introducing an extra implant step called retrograde well implant. High channel concentration also reduces the depletion region formed in source-channel-drain junctions.

Titanium is used to form Silicide in this process. Two phases of Titanium Silicide exit C49 and C54. C49 phase (TiSi) is formed at lower temperature ranging from 500-700°C with a higher resistivity of about 5-7 $\mu\Omega$-cm. The C54 phase (TiSi$_2$) is formed at higher temperature ranging from 700-900°C with lower resistivity of 3-4 $\mu\Omega$-cm. 2.27 nm of Si is consumed per nm of Titanium deposited [4].

**2.4.2 ELECTRICAL PARAMETERS**

Below is a table of comparison between NTRS electrical parameters with the simulated NMOS and PMOS devices. It is important to scale down the voltage and current limits on smaller devices to avoid short channel behavior.

<table>
<thead>
<tr>
<th>ELECTRICAL PARAMETERS</th>
<th>NTRS (130nm) NMOS/PMOS</th>
<th>NMOS (125nm) Simulation Results</th>
<th>PMOS (125nm) Simulation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive current ($\mu$Amps/\mu$m$)</td>
<td>600/280</td>
<td>330</td>
<td>70</td>
</tr>
<tr>
<td>Leakage current ($n$Amps/\mu$m$)</td>
<td>3</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Drain voltage (V)</td>
<td>1.2-1.5</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Threshold voltage (V)</td>
<td>0.4</td>
<td>0.4794</td>
<td>-0.58</td>
</tr>
<tr>
<td>DIBL</td>
<td>&lt;100 mV/V</td>
<td>&lt;100 mV/V</td>
<td>&lt;100 mV/V</td>
</tr>
<tr>
<td>Sub-threshold slope</td>
<td>85 mV/decade</td>
<td>108 mV/decade</td>
<td>96 mV/decade</td>
</tr>
</tbody>
</table>
MOSFET Background

The drive current is compensated because of higher doping concentration in the well. The leakage is slightly high in shorter devices compared to longer channel device. **Drain bias is reduced to avoid gate oxide breakdown** and also to avoid short channel effects like DIBL, sub-threshold slope, channel length modulation and punch-through. Threshold voltage of the device should be designed to be lower than the gate bias. The sub-threshold slope (SS) is parameter which mostly depends on the temperature of operation, oxide capacitance and channel depletion capacitance. SS is an electrical parameter which is a measure of leakage current; off state performance of the device.

### 2.4.3 DESIGN PARAMETERS

Mathematical parameters are used to design the source drain and well doping concentration and control the thickness of the depletion width between source, channel and drain junctions.

#### 2.4.3.1 FIRST ORDER CALCULATION DESIGN

**2.4.3.1.1 BUILT IN POTENTIAL**

Internal potential across the p-n junctions in PMOS or NMOS in thermal equilibrium is given by

\[
\Phi_{bi} = \frac{kT}{q} \ln\left(\frac{N_a N_d}{n_i^2}\right)
\]

where \( k \) is the Boltzmann constant, \( q \) is the charge of an electron, \( T \) is the temperature, \( N_a \) is the acceptor concentration, \( N_d \) is donor concentration and \( n_i \) is the intrinsic carrier concentration.

The values of \( N_a \) and \( N_d \) used in this calculation are obtained from simulation data.
MOSFET Background

NMOS:

\[ \phi_{bi} = 0.026 \ln \left( \frac{44 \times 10^{37}}{2.1025 \times 10^{20}} \right) = 1.0925 \text{ eV} \]

PMOS:

\[ \phi_{bi} = 0.026 \ln \left( \frac{3.567 \times 10^{37}}{2.1025 \times 10^{20}} \right) = 1.0275 \text{ eV} \]

2.4.3.1.2 WIDTH OF THE SPACE CHARGE REGIONS

Width of the depletion layer depends on the concentration of impurities in N-side and P-side of the PN junction. The equations for calculating the width in P-side and N-side is given by

P-SIDE (P-Channel NMOS transistor)

Equation [10] \[ W_{Dp} = \sqrt{\frac{2\varepsilon_s(\phi_{bi}+V_R)}{q\frac{N_D}{N_A(N_A+N_D)}}} \]

N-SIDE (N-channel PMOS transistor)

Equation [11] \[ W_{Dn} = \sqrt{\frac{2\varepsilon_s(\phi_{bi}+V_R)}{q\frac{N_A}{N_A(N_A+N_D)}}} \]

Where, \( \varepsilon_o \) is the permittivity of free space, \( \varepsilon_{si} \) is the permittivity of silicon, \( \phi_{bi} \) is the builtin potential at equilibrium, \( q \) is the charge of an electron, \( N_A \) is acceptor concentration and \( N_D \) is the donor concentration.
MOSFET Background

**NMOS:**

The depletion width in the channel region is much greater than in the n+ source drain regions because the depletion width is inversely proportional to concentration of atoms. Concentration of p-type impurity is less than the concentration of n-type impurities in n+ source drain regions. Hence the formula above allows calculating the depletion width in p region and n-region individually.

\[
W_{Dp}(0V) = \sqrt{\frac{2 \times 11.7 \times 8.854 \times 10^{-14} \times 1.0925}{1.6 \times 10^{-19}}} \left( \frac{8 \times 10^{19}}{5.5 \times 10^{18}(5.5 \times 10^{18} + 8 \times 10^{19})} \right) = 16.3 \text{ nm}
\]

\[
W_{Dp}(1.2V) = \sqrt{\frac{2 \times 11.7 \times 8.854 \times 10^{-14} \times (1.0925 + 1.2)}{1.6 \times 10^{-19}}} \left( \frac{8 \times 10^{19}}{5.5 \times 10^{18}(5.5 \times 10^{18} + 8 \times 10^{19})} \right) = 24.7 \text{ nm}
\]

**PMOS:**

\[
W_{Dn}(0V) = \sqrt{\frac{2 \times 11.7 \times 8.854 \times 10^{-14} \times 1.0275}{1.6 \times 10^{-19}}} \left( \frac{1.5 \times 10^{19}}{5 \times 10^{18}(5 \times 10^{18} + 1.5 \times 10^{19})} \right)
\]

\[W_{Dn}(0V) \text{ Source} = 14.56 \text{ nm}\]
MOSFET Background

\[ W_{Dn}(1.2V) = \sqrt{\frac{2 \times 11.7 \times 8.854 \times 10^{-14} \times (1.0275 + 1.2)}{1.6 \times 10^{-19}} \left( \frac{1.5 \times 10^{19}}{5 \times 10^{18}(5 \times 10^{18} + 1.5 \times 10^{19})} \right)} \]

\[ W_{Dn}(1.2V)_{\text{Drain}} = 21.10 \text{ nm} \]

2.4.3.1.3 EFFECTIVE PHYSICAL GATE LENGTH

Effective gate length is extracted by calculating the depletion width at the source end and at the drain end. These values of depletion widths are subtracted from the mask length.

**Equation [12]** \[ L_{\text{eff}} = L_{\text{mask}} - W(0V)_{SC} - W(1.2V)_{SC} \]

**NMOS:**

\[ L_{\text{eff}} = 150 \text{ nm} - 16.3 \text{ nm} - 24.7 \text{ nm} = 109 \text{ nm} \]

**PMOS:**

\[ L_{\text{eff}} = 150 \text{ nm} - 14.56 \text{ nm} - 21.10 \text{ nm} = 114 \text{ nm} \]

**NOTE:**

These values of effective gate length do not include the lateral diffusion of impurities in silicon after anneal.

Effective electrical channel length is much smaller than the calculated lengths using first order calculations.
MOSFET Background

2.4.3.2 ATHENA AND ATLAS

ATHENA is a modeling tool used to create device structure. Simulations involving dopant and/or its diffusion are very critical for simulation accuracy appropriate model is used to obtain high accuracy. Process steps where correct choices of models are vital include implantation, diffusion, rapid thermal annealing (RTA) and oxidation. These process models are discussed in detail in the Chapter 2.

ATLAS is a simulation tool used to extract the electrical characteristics of the modeled device. It has Shockley read hall, recombination, generation, avalanche breakdown and mobility models to get accurate electrical characteristics similar/comparable to fabricated characteristics. These are discussed in detail in Chapter2.
CHAPTER 3

DEVICE MODELING AND SIMULATIONS

3.1 ATHENA

ATHENA is a 1D, 2D and 3D process simulator used to model semiconductor devices. It is a very effective tool which replaces costly experiments on real time fabrication of devices. Each and every fabrication steps can be modeled in ATHENA which includes oxidation, deposition, ion implantation, conventional furnace annealing, spike anneal (RTA), geometric etches and lithography.

DECKBUILD is the interactive graphic environment used for generating input files for process and devices simulation [5]. The basic operations required to create input files are to develop good simulation grid, perform conformal deposition and geometric etches, saving and loading structure information and interfacing with different simulators like ATLAS.

3.2 GRID STRUCTURE - An initial structure is created using command window in DECKBUILD and select mesh define to create a suitable mesh for the device.

![Grid structure of silicon on which device is modeled.](image)

Fig. 13: Grid structure of silicon on which device is modeled.
Device Models and Simulations

The initial grid for the 100nm CMOS device simulation is as seen in Fig. 13. The mesh grids are very fine in the channel region where the minute of details is important since the gate length is 125nm. And the grid is coarse in the bulk of the silicon.

```
# Grid parameter definition for both x and y co-ordinates
line x loc=0.00 spac=0.1
line x loc=0.8 spac=0.01
line x loc=1.0 spac=0.01
#
line y loc=0.00 spac=0.001
line y loc=0.1 spac=0.01
line y loc=0.2 spac=0.1
line y loc=0.5 spac=0.1
#
struct mirror right
```

3.3 INITIAL SUBSTRATE

Initializing the substrate with right resistivity, material, type and orientation is very important to match the actual fabricated wafers.

![Graph](image)

Fig. 14: Starting P-type silicon substrate with <100> orientation.

The device is modeled on a P-type wafer, with background wafer concentration defined by the resistivity of the wafer. To match the actual fabricated wafers’ resistivity; device is modeled on a silicon substrate with resistivity 20 Ω-cm.
Device Models and Simulations

Orientation of the silicon substrate is defined as <100>. These parameters were selected in a mesh initialize menu under command window.

# Mesh initialize
init silicon boron resistivity=20 orientation=100 two.d
#
struct mirror right

3.4 ION IMPLANTATION MODELS

Ion implantation is one of the key methods to dope the semiconductor with desired impurities. It is one of the critical steps since modern MOSFET technologies have very shallow junctions, tilted implants and high doses. So, the simulation accuracy for implant process is very critical. ATHENA has four different implant models. The advantage of using one model over the other is addressed. These models are Gaussian, Pearson model, SIMS Verified Dual Pearson (SVDP) model and Monte Carlo Binary Collision Approximation (BCA) model.

3.4.1 GAUSSIAN MODEL

Gaussian model is the simplest implant model used for creating profiles on 1D structure. The mathematical model for Gaussian implant is given by [5]

\[
C(x) = \frac{\varphi}{\sqrt{2\pi\Delta R_p}} \exp\left(-\frac{(x-R_p)^2}{2\Delta R_p^2}\right)
\]

Where \(\varphi\) is dose/cm\(^2\), \(R_p\) is projected range, \(\Delta R_p\) is straggled projected range.

Gaussian model is not adequate to calculate real implant profiles, because real implant profiles are asymmetrical.
**3.4.2 PEARSON MODEL**

Pearson IV functions are widely used models for calculating asymmetrical implant profiles. This model is used to get longitudinal implant profiles. This function solves series of differential curves and therefore it is represented by [5],

\[
\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b_0 + b_1 x + b_2 x^2}
\]

Where \(f(x)\) is the frequency function, \(a\), \(b_0\), \(b_1\) and \(b_2\) are moments of \(f(x)\). Results obtained using single Pearson model is better than the results obtained using Gaussian model. But this model excludes the effect of channelling. Athena developed another model to address this issue called Dual Pearson model.

**3.4.3 DUAL PEARSON MODEL**

Since channelling is the main issue in some of the implant profiles, Dual Pearson model addressed this issue. It has a linear combination of two Pearson functions. First Pearson function represents random scattering part and the second function represents channelling tail region [5].

\[
C(x) = \phi_1 f_1(x) + \phi_2 f_2(x)
\]

\(\phi_1\) is the dose for first Pearson function \(f_1(x)\), \(\phi_2\) is the second Pearson function for \(f_2(x)\).

\[
C(x) = \phi [R f_1(x) + (1 - R) f_2(x)]
\]

Where, \(R\) is the dose ratio and \(\phi\) is the total dose.

Dual Pearson implant needs nine model parameters four for each Pearson function and dose ratio \(R\). This model does not work well for implants on amorphous material.
Device Models and Simulations

So more advanced model is developed called SVDP (SIMS Verified Dual Pearson) model.

### 3.4.4 SIMS Verified Dual Pearson (SVDP)

SVDP is the default model used by ATHENA. It is accurate if the implant energies and doses used are within the range specified in Table 3 obtained from University of Texas Austin, where high level experiments were conducted on B, BF$_2$, P and As impurities.

Table 3: Accuracy of SVDP implant model ranges [5].

<table>
<thead>
<tr>
<th>Ions</th>
<th>Energy (keV)</th>
<th>Dose (cm$^{-2}$)</th>
<th>Tilt Angle ($\theta$)</th>
<th>Rotation Angle ($\phi$)</th>
<th>Screen Oxide (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1-100$^a$</td>
<td>$10^{13}$-8x10$^{15}$</td>
<td>0-10</td>
<td>0-360</td>
<td>native oxide-500$^b$</td>
</tr>
<tr>
<td>BF$_2$</td>
<td>1-80$^c$</td>
<td>$10^{13}$-8x10$^{15}$</td>
<td>0-10</td>
<td>0-360</td>
<td>native oxide</td>
</tr>
<tr>
<td>P</td>
<td>12-200$^d$</td>
<td>$10^{13}$-8x10$^{15}$</td>
<td>0-10</td>
<td>0-360</td>
<td>native oxide</td>
</tr>
<tr>
<td>As</td>
<td>51-200$^e$</td>
<td>$10^{13}$-8x10$^{15}$</td>
<td>0-10</td>
<td>0-360</td>
<td>native oxide</td>
</tr>
</tbody>
</table>

This model do not work well for multi layer implants, because nuclear and electronic stopping powers are different for different materials. So several scaling methods were developed in ATHENA like dose match parameters, scaling parameters for different materials, Gaussian and non Gaussian lateral diffusion parameters. But, these parameters did not give accurate results.

A different model was developed by SILVACO to address abovementioned issues called Monte Carlo implant model.

#SVDP Implant statement
implant boron dose=2.0e15 energy=75 smooth=0.50 rotation=45 crystal
3.4.5 MONTE CARLO IMPLANT MODEL

Monte Carlo implant model gives close to accurate results on multi layer and non planar structures. Multi layer implant accuracy using this model is comparable to single layer structure. This model is based on binary collision approximation. This model includes the nuclear stopping power and electronic stopping power mathematical model for impurities implanted and the crystal structure. Implants on amorphous, crystalline structures can be modeled with accuracy since the development of Monte Carlo implant model.

3.4.5.1 Nuclear Stopping

The ions implanted not only interact with the crystal atoms but also with the electrons [5]. In this model it is assumed that ions move in a straight line segment after deflection and ions loses energy through this elastic scattering. The scattering angles of projectiles and recoils are given by [5],

\[ \tan \theta_1 = \frac{Af \sin \theta}{(1+A f \cos \theta)} \]
\[ \tan \theta_2 = \frac{f \sin \theta}{(1 + f \cos \theta)} \]

Where,

\[ f = \sqrt{1 - \frac{Q}{E_r}} \]
Device Models and Simulations

$Q$ is the energy lost by electron excitation, $A$ is the ratio of the mass of the target atom (Si atom in this case) by mass of the projectile implant atom (B, BF$_2$, P and As), $\Theta$ is the scattering angle and $E_r$ is the relative kinetic energy.

3.4.5.2 Interatomic Potential

A two body screening Coulomb potentials with screening functions are used by ATHENA. It is a numerical fit solution given by Firsov [6], the analytical form is similar to isolated atom [5].

$$V(r) = \frac{Z_1 Z_2 e^2}{r} \kappa \left( \frac{r}{a_0} \right)$$

Where, $Z_1$ and $Z_2$ are atomic numbers of two atoms, $a_0$ is the screening length and $\kappa$ is the screening function.

3.4.5.3 Electronic Stopping

ATHENA uses two separate mechanisms to model inelastic energy losses; local and non-local. Both models have different energy and special dependencies.

The local inelastic model is based on Firsov model [6], the electronic energy loss per collision in this model is based on average energy of excitation of electron shells and motion and distribution of an atom [5]. And is given by [5].

$$-\Delta E = \frac{0.05973 \left( Z_1 + Z_2 \right)^{5/3} \sqrt{E/M_1}}{(1 + 0.31(Z_1 + Z_2)^{1/3} R_0)^5}$$

Where, $\Delta E$ is the transfer of energy from ion to an atom by electron passage from one particle to another, $Z_1$ and $Z_2$ are atomic number of two atoms, $M_1$ is mass in a.m.u., $E$
Device Models and Simulations

is the energy of moving atom and $R_0$ is distance of the closest approach, which is approximately equal to impact parameter [5].

The non local inelastic model is based on model proposed by Brandt and Kitagawa [7]. In this model the stopping power is represented by, $= - \frac{dE}{dx}$; the medium of an ion is proportional to mean-square effective ion charge. Brandt and Kitagawa derived effective stopping power charge $Z^*_i$ from the ionization state $q$. The fractional effective charge of an ion is given by [5],

\[
\zeta \equiv \frac{Z^*_i}{Z_1} = \left[ \frac{S_q}{S_{q=1}} \right]^{1/2}
\]

Where, $S_{q=1}$ is bare nucleus stopping power, the equation for fractional effective charge of an ion is given by,

\[
\zeta \approx q + C(k_F)(1 - q) \ln(1 + (2\Lambda v_F a_0 v_0)^2)
\]

Where, $q$ is fractional ionization, $a_0$ and $v_0$ is Bohr’s radius and velocity, $k_F$ and $v_F$ is fermi wave vector and velocity, $\Lambda$ is screening radius.

# Monte Carlo implant statement
implant bf2 dose=9.0e14 energy=20 monte n.ion=100000 smooth=0.25 \rotation=45 unit.dam dam.factor=0.05

3.4.5.4 Damage accumulation model

This model is to transform the crystalline substrate to amorphous state after implant. Deposited energy and number of projectiles accounts for the damage done to crystalline surface. This function is represented by [5],
Device Models and Simulations

\textbf{Equation [23]} \quad f(r) = 1 - \exp \left( \frac{\Delta E(r)}{E_c} \right)

Where, \( \Delta E(r) \) is deposited energy per unit volume at grid point \( r \), \( E_c \) is critical energy density representing total energy needed to amorphize the crystal structure.

This energy is given by [5],

\textbf{Equation [24]} \quad E_c(T) = E_{c0} \left( 1 - \exp \left[ \frac{E_0(T-T_\infty)}{2k_B T T_\infty} \right] \right)^{-2}

Where, \( E \) is activation energy, \( k_B \) is Boltzmann constant and \( T_\infty \) is the temperature at which infinite dose is required to amorphize the crystal structure.

\# Monte Carlo implant statement
implant bf2 dose=9.0e14 \textbf{energy}=20 \textit{monte n.ion=100000 smooth=0.25 } \backslash \text{rotation=45}

\textbf{3.5 DEPOSITION MODEL}

Deposit statement is used to deposit a material and the thickness of the material to be deposited is also specified in the same statement.

\# deposit gate oxide (oxynitride)
deposit oxynitride thick=0.003 divisions=5

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig15.png}
\caption{Conformal deposition of gate oxide using deposition model.}
\end{figure}
Device Models and Simulations

Sometimes in critical deposition steps like gate oxide, the deposit layer is divided into number of layers with controlled grid spacing. This step is important to simulate the flow of charge carriers through the dielectric. The division of layers is controlled by DIVISION parameter in deposit statement. Grid of the deposited material is controlled by DY parameter which specifies nominal spacing.

### 3.6 ETCHING MODEL

Etching is an important part of fabrication. But ATHENA considers etching as a geometrical problem and ignores impurity redistribution effects involved during etch process. Etches are assumed to be performed at room temperature.

**Fig. 16: Geometrical etch of Polysilicon in ATHENA.**

Geometrical etches in ATHENA are performed using ETCH statement. The material, geometry (x and y co-ordinates are used to specify geometry) and thickness of the material to be etched are specified in this statement. Fig 16 shows the desired geometrical shape in Polysilicon after using ETCH statement.

```
# etch statement
etch polysilicon start x=0.00 y=-0.25
etch cont x=0.85 y=-0.25
etch cont x=0.85 y=0.00
```
3.7 DIFFUSION MODELS

Diffusion is an important step in modeling of 100 nm devices. All models in ATHENA work on the principle of pair diffusion. Point defect assists the diffusion of impurities thus the term pair diffusion. Point defects can diffuse either as cluster or as a dopant-defect pair, the diffusion properties of point defects can change depending on the type of model selected. So, appropriate model has to be used to accurately model diffusion of source drain extensions and source drain regions; since it is critical step of fabricating 100 nm CMOS devices.

3.7.1 FERMI DIFFUSION MODEL

Fermi diffusion model does not directly represent point defects, they are assumed to be in thermal equilibrium. Since pair diffusivity of point defects and dopants diffusion constitute fermi model, defect population are not in equilibrium throughout the simulation.

Fermi model is the default diffusion model in ATHENA. Dopant diffusion obeys continuity equation of the form

$$\frac{\partial C_A}{\partial t} = \nabla \left[ (D_{AV} + D_{AI}) C_A + \nabla \ln \left( C_A + \left( \frac{n}{n_i} \right)^z \right) \right]$$

Where, $n/n_i$ is the factor that accounts for electric field effect, $z$ is the direction of electrical force (+1 for donors and -1 for acceptors), $D_{AV}$ and $D_{IV}$ are the point defects contributing dopant diffusivity. Total diffusivity depends on both temperature and fermi level. Intrinsic carrier concentration in ATHENA is given by [5],
Device Models and Simulations

\textbf{Equation [26]} \quad n_i = n_{i0} \exp \left( -\frac{n_i E}{kT} \right) T^{n_{ipow}}

Where \( n_{i0}, \ n_{ig} \), and \( T^{n_{ipow}} \) are the material parameters to be specified for diffusion in MATERIAL part.

\textbf{Method fermi} – Diffusion statement used to run fermi model.

\subsection*{3.7.2 IMPURITY SEGREGATION MODEL}

The dopant transport in the interface is also considered and modeled in ATHENA.

The interface can either be solid-solid or gas-solid.

This model is given by first order kinetic model for flux \([5]\),

\textbf{Equation [27]} \quad F_s = h_{12} \left( \frac{C_1}{M_{12}} - C_2 \right)

Where \( C_1 \) and \( C_2 \) are the impurity concentrations at the interface of two materials, \( h_{12} \) is the velocity of impurity transport and \( M_{12} \) is the segregation coefficient given by \([5]\).

\textbf{Equation [28]} \quad M_{12} = \frac{\text{solid solubility of impurity in material 1}}{\text{solid solubility of impurity in material 2}}

\( M_{21} \) is the inverse of \( M_{12} \).

The impurity velocity transport is given by \([5]\),

\textbf{Equation [29]} \quad h_{21} = M_{12} \ h_{12}

Since point defects are considered to be in thermal equilibrium in fermi model, there is no separate segregation model for vacancies and interstitial concentrations.
Device Models and Simulations

3.7.3 TWO DIMENSIONAL MODEL

Two dimensional model is the extended or an improved version of fermi model. This model has separate representations for point defects, so there are three sets of equations for dopant diffusion.

The continuity equation for this model is given by [5],

\[
\text{Equation [30]} \quad \frac{\partial C_A}{\partial t} = \sum_{X=I,V} \nabla \left[ \left( R_{IV} D_{AX} \right) C_A + \nabla \ln \left( Z_A C_A + \frac{C_x}{C_x^*} \left( \frac{n}{n_i} \right) \right) \right]
\]

Where \( C_X \) is the point defect concentration, \( C_X^* \) is the equilibrium concentration of the point defect and \( R_{IV} \) is the contribution of interstitials and vacancies in dopant diffusion and it is also temperature dependent. \( C_X/C_X^* \) is the defect ratio; it is temperature and fermi level dependent.

The continuity equation for interstitial profile is given by [5],

\[
\text{Equation [31]} \quad \frac{\partial C_I}{\partial t} = \nabla \left( -J_I \right) - R + \frac{\partial C_{ET}}{\partial t} + R_{<311>}
\]

Where \( R \) is the recombination rate in the bulk (recombination between interstitials and vacancies), \( J_I \) is the flux which accounts for electric field distribution, \( \frac{\partial C_{ET}}{\partial t} \) is the capture and emission of interstitial traps [5].

The continuity equation for vacancies is given by,

\[
\text{Equation [32]} \quad \frac{\partial C_V}{\partial t} = \nabla \left( -J_V \right) - R
\]

Where \( R \) is the recombination rate in the bulk (recombination between interstitials and vacancies), \( J_I \) is the flux for vacancies which accounts for electric field distribution.
Device Models and Simulations

**Method Two.Dim** – Diffusion statement used to run Two dimensional model.

### 3.7.4 FULLY COUPLED MODEL

Fully coupled model is more advanced model than two dimensional model where two way coupling between diffusion of dopants and point defects is implemented. It is done by adding dopant-defect flux terms to defect part of the equation. So the continuity equation for both interstitials and vacancies is given by [5],

\[
\frac{\partial C_V}{\partial t} = \nabla (-J_V - \sum_{A,C} J_{AV}^C) - R
\]

**Equation [33]**

\[
\frac{\partial C_I}{\partial t} = \nabla (-J_I - \sum_{A,C} J_{AV}^C) - R + \frac{\partial C_{ET}}{\partial t} - \frac{\partial C_{<311>}}{\partial t}
\]

**Equation [34]**

This model is very much similar to two dimensional model except that the summation runs over all dopants and pair charge states. This model works well for high dopant and high implant damage conditions unlike two dimensional model.

**Method full.cpl** – Diffusion statement used to run fully coupled model.

**Method full.cpl cluster.dam** – Diffusion statement with fully coupled and cluster.dam is used to diffuse point defects implanted and enhance the diffusion due to the effects of transient enhanced diffusion (TED) which is prominent in low temperature; using Monte Carlo implants.
Device Models and Simulations

3.7.5 RAPID THERMAL ANNEALING

ATHENA can simulate RTA using two dimensional and fully coupled diffusion models. RTA is a short thermal process where low temperature ranges (800°C to 1000°C) is avoided to obtain shallow junctions. Temperature ramp up rate is very high. In this work RTA is used to avoid the enhanced diffusion due to point defects at source drain anneal. It is also used to form metal Silicides to form good contacts.

Table 4: Duration of TED at various temperature ranges [5].

<table>
<thead>
<tr>
<th>Anneal Temperature (°C)</th>
<th>Time for completion of 95% TED</th>
</tr>
</thead>
<tbody>
<tr>
<td>600</td>
<td>390 hours</td>
</tr>
<tr>
<td>700</td>
<td>3.3 hours</td>
</tr>
<tr>
<td>750</td>
<td>30 min</td>
</tr>
<tr>
<td>800</td>
<td>3.7 min</td>
</tr>
<tr>
<td>850</td>
<td>43 sec</td>
</tr>
<tr>
<td>900</td>
<td>8.3 sec</td>
</tr>
<tr>
<td>950</td>
<td>1.9 sec</td>
</tr>
<tr>
<td>1000</td>
<td>0.48 sec</td>
</tr>
<tr>
<td>1050</td>
<td>0.13 sec</td>
</tr>
</tbody>
</table>

Table 4 is approximate time for completion of 95% of TED that ATHENA uses at different temperatures.
Comparison between FERMI, TWO DIMENSIONAL and FULLY COUPLED models on a plane surface using conventional furnace recipe:

A model comparison between Fermi, Two Dimensional and Fully Coupled diffusion is performed using conventional furnace anneal recipes. Monte Carlo implant is used with a phosphorus dose of $2 \times 10^{15} \text{ cm}^{-2}$ and energy of 30 $\text{KeV}$. Impurities are implanted on a planar surface with a screening oxide thickness of 250 $\text{Å}$. Soak time and temperature used in the recipe is 4 hours and 1000 $\text{°C}$ respectively. The recipe also includes ramp up from 800°C to 1000°C for 20 minute and ramp down from 1000°C to 800°C for 20 minutes.

From observing Fig. 17, it can be concluded that the fermi model works well for deeper junctions with accuracy of junction depth matching fully coupled and two dimensional models. But the concentration of impurities in the surface of the silicon is slightly higher than two dimensional and fully coupled model indicating...
Device Models and Simulations

the presence of impurity segregation in two dimensional and fully coupled models.

**Comparison between FERMI, TWO DIMENSIONAL and FULLY COUPLED models on a planar surface using RTA recipe:**

![RTA Comparison Between FERMI/TWO_DIM/FULLY COUPLED Models](image)

Fig. 18: RTA diffusion comparison between fermi, two dimensional and fully coupled models.

RTA diffusion results using Fermi, Two Dimensional and Fully Coupled models are compared. Implant impurity used is Phosphorus and implant dose is $2 \times 10^{15}$ $cm^{-2}$. The implant energy is 30 $KeV$ and impurities are implanted through 250 $Å$ screen oxide. The RTA diffusion recipe has 3 seconds ramp up from 650 to 1100 °C, followed by 5 seconds soak at 1100 °C in Nitrogen ambient and ramp down from 1100 °C to 650 °C in 6 seconds.

From observing Fig. 18, it is noticed that fermi model is not accurate giving smaller junction depth compared to two dimensional and fully coupled models in which diffusion of point defects are explicitly included, unlike in fermi model where point defect concentration is assumed to be in equilibrium.
Device Models and Simulations

The concentration of the impurity is higher \((3.2 \times 10^{19} \text{ cm}^{-3})\) in fermi model at the surface than in fully coupled model or two dimensional model \((1.3 \times 10^{19} \text{ cm}^{-3})\).

3.8 ATLAS

ATLAS is electrical device simulation software used for characterizing 2D and 3D models. It can simulate the electrical characteristics for all semiconductor devices structures. It provides insight into the physics of the operation of the device.

Some of the Physics models used to simulate MOSFETS are CVT and SRH.

3.8.1 LOMABARDI CVT MODEL (CVT)

This model is inversion layer model which overrides any other mobility model [8]. This model is selected by typing CVT in the model statement. The parameters that are associated in this mobility model are transverse electric field, doping and temperature dependence. This model also accounts for mobility degradation in the inversion layer due to surface scattering in the interface of insulator-semiconductor.

3.8.2 SHOCKLEY READ HALL MODEL (SRH)

Shockley Read Hall is used to model the recombination effects in a device model [8]. This model simulates the leakage current due to thermal effects and also it is useful to simulate the presence of interface trap charges. This model also considers electron and holes carrier lifetime and lattice temperature to simulate leakage currents.
CHAPTER 4

UNIT PROCESS IMPROVEMENTS

4.1 WELL IMPLANT, RETROGRADE WELL IMPLANT AND SOURCE DRAIN EXTENSION IMPLANT DOSE

As device geometries scale, there is an increasing requirement for high control of all process steps related to doping.

Well implant dose and retrograde implant dose give the concentration of impurities in the surface of the silicon. The design of right dose for well implant, retrograde well implant and source drain extension implant is critical to obtain desired/designed threshold voltage for the modeled device and to control the lateral diffusion of impurities during source drain anneal.

One important reason in high doping concentration of source drain extensions is to lower the impact of series resistance (highly doped regions are less resistive) to better the on state performance of the device.

The N+ and P+ polysilicon gates for NMOS and PMOS are highly doped to suppress the effects of gate depletion.

4.2 GATE OXIDE GROWTH

The thickness of the gate oxide is scaled down along with length of the gate. For gate lengths ~130 nm NTRS road map requires a 30 Å thick gate oxide. Thinner gate oxide gives gate more control over the channel region.
Unit Process Improvements

With thin gate oxide, the capacitance across the channel region is increased and hence the change in charge per change in voltage is large. This results in faster switching speed from on state to off state and vice versa.

At this point the use of high k dielectric like hafnium dioxide is very important. Gate leakage is significant on devices which uses SiO\textsubscript{2} as gate oxide with nitrogen incorporation. Gate leakage was the major issue for the fabricated device failures. These failures are discussed in detail in Chapter 5.

**4.3 DOUBLE EXPOSURE**

Lithography is the biggest hurdle (in process point of view) to attain the mask length of 150 nm. The minimum resolution that ASML stepper at RIT has is 0.35 µm. So to obtain minimum feature size of 150 nm, double exposure technique is used.

Double exposure has two lithography passes. In the first pass, the wafers are aligned with alignment marks from previous lithography and exposed die after die like every other exposure. In the second pass the wafer stage is shifted in Y axis and this shift is set by user according to his process requirements.

In this case, 1 micron features were double exposed by shifting the wafer stage by 0.8 µm in the second pass. Resulting feature size obtained is 0.15 µm and ~0.05 µm are a part of losses due to scattering or reflection of UV rays.

**4.4 POLYSILICON AND NITRIDE ETCH (RIE)**

**POLYSILICON** is etched on Drytek Quad; it is RIE tools used to obtain anisotropic etch profiles. That is, the vertical etch rate is greater than horizontal avoiding undercuts (good aspect ratio).
The gases used for etching polysilicon are SF$_6$ (Sulphur Hexafluoride), CHF$_3$ (Tri-Fluoro Methane), O$_2$ (Oxygen) and the flow rates are 30 sccm, 30 sccm and 5 sccm respectively. Fluoride ions obtained from SF$_6$ and CHF$_3$ are bombarded on the surface of the exposed polysilicon; bombardment of these ions provides physical damage and fluoride ions react with polysilicon and outgas. The carbon and hydrogen in Tri-Fluoro Methane helps in forming a polymer on the walls of the polysilicon protecting the walls from under etch. Oxygen gas is added in the recipe to remove the polymer formed on side walls and improve the selectivity of etch between polysilicon and underlying oxide. The etch rate of polysilicon after etch is found to be $1150 \text{ Å/min}$ and etch rate of oxide is $200 \text{ Å/min}$. 

Fig. 19: SEM image of Polysilicon gate.
Unit Process Improvements

Polysilicon starts etching outside to inside of the wafer on DryTek Quad. Plasma process can be watched through the side window for the completion of polysilicon etch and the change in the plasma colour is observed. Concentric circle with change in plasma colour starting from the edge to centre of the wafer indicates the completions of polysilicon etch.

The walls of the polysilicon is damaged in the process, it is repaired by a thermal step called Poly-Reoxidation.

**NITRIDE**

Nitride is also etched on Drytek Quad using RIE technique. The gas used for nitride etch is SF$_6$ at a chamber pressure of 40 mTorr and RF power used to etch nitride is 250 Watts. Since SF$_6$ is the only gas contributing Fluoride ions to etch nitride and
Unit Process Improvements

there is no gas used that provide C and H atoms to improve anisotropy; isotropic profile is observed causing the spacer to be overetched. Overetch is characterized by measuring the width of spacer and comparing it with deposited thickness.

The reason behind thicker polysilicon deposition and wider LDD used in this process is because phosphorus is used as an impurity for NMOS source drain region. So, more room for space is required to avoid the diffusion of phosphorus into the channel region. Phosphorus is used in this process to make the device more manufacturing capable at RIT.

The etch rate of nitride using this process condition is calculated to be 1120 Å/min.

4.5 LOW RESISTANCE SOURCE DRAIN CONTACTS

The source drain regions are much deeper and these regions are used for contacts. These regions are highly doped to reduce the contact resistance. High concentration silicon surface in conjunction with Silicide formation makes, the contacts more ohmic with very less forward voltage drop.

These junctions are deeper than source drain extensions because Silicide consumes part of silicon when they undergo thermal process. Junction depth of source drain regions after anneal in this process is ~100 nm to 120 nm.

The energy of implant used to attain above mentioned junction depth is 25 KeV to 30 KeV depending upon the impurities implanted.

4.6 SPIKE ANNEAL

The control over dopant impurity diffusion using a thermal process is important. There are multiple reasons why spike annealing is preferred over conventional
Unit Process Improvements

furnace annealing. One of those is to avoid anneal at low temperatures. At the temperature range of 750 to 1000ºC the TED is significant and enhances dopant diffusion.

To avoid these temperature ranges, RTP tool is used to spike the temperature from 700ºC to 1050ºC in one second. The obtained profiles after this process is ultra shallow compared to junction profile obtained from furnace anneal.

Since the junctions formed are ultra shallow, the resulting sheet resistance is also high. This contributes to a degraded on state performance of the device. Studies have shown that spike annealing to 1050ºC for couple of seconds is enough to 100% activate the implanted dopant impurities.

Second of the reason is thermal budget reduction by using spike annealing over furnace annealing.

4.7 TITANIUM SILICIDE

Since the source drain junctions are very shallow, a thin film of titanium is deposited to form very thin layer of Silicide. It is important not to consume more than half of the junction depth of source drain region, because it increases the off state current in the device.

Titanium is sputtered using sputtering technique and a recipe was developed to deposit 300Å of Titanium. 4 inch target is used to sputter Titanium and power distributed on the target is 350 Watts, deposition pressure is 1x10⁻⁶ Torr. Argon gas is used in the sputtering process since Ar is huge atom. Titanium is sputtered for 4 minutes in the above mentioned process condition.
Unit Process Improvements

The wafers are now annealed at 650°C for 2 seconds to form TiSi (Titanium Silicide). TiSi formed after this process is known as C49 phase. The resistance of contacts is high in this phase of Silicide. So, after blanket etch of unreacted Ti in hot sulphuric acid bath, the wafers are annealed again at 715°C to form TiSi$_2$ (C54 phase) which is less resistive than C49 phase. The barrier height is reduced after second thermal step (C54 phase) than it is in C49 phase allowing more charges to flow over the barrier upon metal semiconductor contact.
CHAPTER 5

FABRICATION OF 100 nm (Leffective) CMOS DEVICES

5.1 INTRODUCTION

Fabrication of 100nm CMOS devices are carried out in SMFL (RIT’s clean room facility for research and development). As a part of process development, recessed oxide growth is used for isolation. Heavy impurity ions like Arsenic and BF$_2$ are used and low energy implant is used for shallow junctions. Dual work-function degenerately doped N-type and P-type polysilicon gates are used for NMOS and PMOS respectively. Retrograde implants are done to avoid short channel effects. Spike annealing is used to limit thermal budget and to anneal implant damage and for minimal diffusion. Double exposure technique is used to obtain the feature size of 150nm. SALICIDE technique is used to get source drain ohmic contact.

A hand carry LOT (F140214) with three device wafers and three control wafers are processed. MESA is used to keep track of process on device wafers.

The mask includes other structures like CBKRs, Vander Paws, Ring oscillators and some MEMS sensors.

This chapter will also include cross-sections obtained from ATHENA which pictorially demonstrates each process step. Pictures of all process steps are included except for lithography.

These pictures provide detailed information of ion implantation and diffusion steps involved in the whole process.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.2 STARTING WAFER

P-type starting wafer is used with a resistivity of 20 $\Omega$-cm. The thickness of the starting wafer is 650 $\mu$m. The crystal orientation of these three device wafers and three control wafers is $<100>$. One side of the wafer is polished and the other side is not. Polished side of the wafer is used to fabricate CMOS devices.

5.3 PAD OXIDE GROWTH

Thermal oxide is grown on starting wafers on BRUCE Furnace Tube 4 dedicated for dry oxide growth. Recipe 250 is used to grow pad oxide, the expected thickness of oxide from this recipe is 500 Å. The recipe details is as follows, the tube is warmed up to 800°C using recipe 888. Then the recipe 215 is started; the three device wafers are loaded into the furnace with an estimated boat in time of 10 minutes. The recipe has temperature stability time for 30 minutes, where the tube is stabilized at 800°C in nitrogen ambient. The next step is 20 minutes ramp up, the temperature is ramped up from 800°C to 1000°C.

![Initial P-type substrate with 500Å pad oxide grown on it.](image)

Fig. 21: Initial P-type substrate with 500Å pad oxide grown on it.
Fabrication of 100 \( nm \) \( L_{\text{eff}} \) CMOS Devices

The wafers are then soaked for 56 minutes in a dry \( O_2 \) ambient at 1000\(^{\circ}\)C. It is followed by 5 minutes anneal step at 1000\(^{\circ}\)C in nitrogen ambient. The furnace is then ramped down from 1000\(^{\circ}\)C to 800\(^{\circ}\)C in 40 minutes. The boat out time is 15 minutes.

The wafers are then taken out and Nanometrics Spectrometer is used to measure oxide using 81 points thermal oxide recipe. The average thickness measured is \( 438.45 \) Å. Pad oxide acts as a stress relief layer for the following nitride deposition step.

**5.4 NITRIDE DEPOSITION**

Nitride is deposited using LPCVD technique. It is deposited on lower tube of ASM LPCVD. The temperature and gases to be used for deposition is initialized running a cassette on the tool.

![CVD Nitride deposition](image)

Fig. 22: CVD Nitride deposition.

Factory Nitride 810 recipe is used for deposition. The deposition temperature is 810\(^{\circ}\)C, two gases Ammonia (\( \text{NH}_3 \)-150 sccm) and Dichlorosilane (\( \text{SiH}_2\text{Cl}_2 \)-60 sccm) is used for nitride deposition. Deposition time is calculated using the most recent nitride deposition rate noted on the log sheet. In this case, nitride is deposited for 25 minutes with a base pressure of 300 \( m\text{Torr} \).
Fabrication of 100 nm $L_{eff}$ CMOS Devices

The wafers look golden yellow after the run, the thickness of the nitride layer deposited is measured using spectrophotometer using nitride on oxide recipe. The average thickness measured is 1399.7Å.

The time, temperature and deposition rate is updated on the log sheet.

5.5 FIRST LEVEL PHOTO (RECESSED OXIDE)

Following nitride deposition, the field areas are defined using STI mask.

The wafers are coated with OiR 620-10 photoresist on SSI track. Before coating the wafers with resist HMDS priming is done to improve adhesion of resist coated with underlying layer. Recipe COAT.RCP is used to coat the wafers, the resist coated on the wafers is 1.0 µm thick. And the wafers with resist are baked at 110°C.

The wafers are then exposed on i-line ASML stepper using ADV_STI mask. The job used on the stepper is fac_adv_cmos, the exposure energy used is 250 mJ/cm². The illumination mode used is conventional, with numerical aperture equals 0.68.

Exposed wafers are then developed using the develop track on SSI. Photoresist is sensitive to UV light when exposed and developed using CD-26 DEVELOPER. The exposed areas dissolve and mask is imprinted on the wafer. The wafers are baked two times post expose and post develop. Post exposure bake is for 60 seconds at 110°C (Soft bake) and pre exposure bake is for 60 seconds at 140°C (Hard bake).

5.6 NITRIDE ETCH

Plasma etch technique is used to etch nitride in exposed field areas. Nitride is etched on LAM 490 which is a dry etch tool. This tool provides an advantage of optical end point detection. The colour of the plasma is monitored while etching a material
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

(Nitride in this case). When it starts etching underlying material then the colour of the plasma changes (thermal oxide in this case) and indicates that it’s etching a different material.

Fig. 23: Plot of light sensitivity vs time in seconds.

The figure above shows the plot of light sensitivity (Y-Axis) vs time (X-Axis). The drop in the sensitivity of the light is observed when oxide starts etching.

Fig. 24: Nitride plasma etch.

Etchant gases used in this tool is SF$_6$ (200 sccm) and a base pressure of 200 mTorr.

Fluorine ions are used for etching nitride.

FNIT1500.RCP is used to etch nitride at this step. This recipe has three different steps, first is a time delay step for 60 seconds with no RF power, second is nitride etch
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

step for 150 seconds with RF power 125W and third is 40% over etch with the same RF power (125W).

Over etch is to ensure nitride is etched throughout the wafer. And underlying pad oxide thickness is measured after etch. Thickness measured is in the range of 245Å to 314Å.

5.7 PHOTORESIST STRIP

Following nitride etch, photoresist is stripped on GASONICS AURA 1000. Photoresist is stripped in oxygen plasma, using recipe FF. FF is 198 seconds strip recipe with two heat steps, one at the beginning of the recipe and one after 99 seconds.

5.8 RCA CLEAN

Wafers from GASONICS are then moved to RCA clean for surface cleaning from resist residue contaminants before furnace anneal.

RCA clean is done to remove both organic and metal contaminants. There are two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

There is a short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 to remove chemically grown native oxide from SC1 bath.

The process flow of the RCA clean is as shown in the figure below,

![RCA clean process flow diagram]

**Fig. 25: RCA clean process flow.**

### 5.9 FIRST OXIDE GROWTH (RECESSED OXIDE GROWTH)

Field oxide growth is a wet oxide growth. Bruce Furnace Tube 1 is used for wet oxide growth. The tube is initially heated using recipe 888 up to 800°C and then recipe 336 is loaded into the furnace with device wafers in the boat.

First step in the recipe 336 is to stabilise at 800°C for 27 minutes in nitrogen ambient. It is followed by a ramp up step from 800°C to 1000°C for 20 minutes in nitrogen ambient. The soak time is 55 minutes at 1000°C in wet O$_2$ ambient. The recipe has a ramp down time of 35 minutes from 1000°C to 800°C.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Oxide is measured on a nanospec using thick oxide on silicon recipe. Oxide is measured on 5 different points on the wafer; top, bottom, left, right and centre. Oxide is measured is $3514\,\text{Å}$, $3688\,\text{Å}$, $3441\,\text{Å}$, $3416\,\text{Å}$ and $3451\,\text{Å}$.

**5.10 OXIDE ETCH (WET ETCH)**

Oxide is etched on a wet chemical bench. It is etched in 10:1 ($\text{H}_2\text{O}:\text{HF}$) HF bath called BOE. Etch rate of thermally grown oxide in this bath is $586\,\text{Å/min}$ at room temperature. Etch time is 6 minutes 30 seconds; 30 seconds over-etch to get undercut.

After oxide etch, the wafers are rinsed in DI water and dried using SRD. The thickness of the oxide measured on field areas read less than 20Å.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.11 SECOND OXIDE GROWTH (RECESSED OXIDE)

Thermal oxide is grown again using recipe 336 and the target thickness is 3500Å. This run is also carried out on Bruce furnace 1 dedicated for wet oxide growths.

![Second field oxide re-growth](image)

Fig. 28: Second field oxide re-growth.

As seen in the figure, oxide re-growth gives a planar surface unlike LOCOS. But bird’s beak cannot be eliminated by this process. Recessed oxide is the method used for isolation in the absence of CMP for STI.

After second oxide growth the oxide is measured on nanospec and the readings are 3549Å, 3539Å, 3524Å, 3507Å and 3539Å.

5.12 NITRIDE ETCH (WET ETCH)

Nitride is etched in hot phosphoric acid bath and is etched at 165°C. Etch rate of nitride at this temperature is 80Å/min. Nitride is etched for 45 minutes and boat should be frequently stirred to provide agitation and temperature of the bath should be the same throughout the etch process to be uniform.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

Fig. 29: Cross sections of recessed oxide isolation after nitride etch.

Selectivity is good using hot phosphoric acid etch; protecting the underlying pad oxide from overetch.

5.13 SECOND LEVEL PHOTO (NWELL)

The second photo step is for n-well implant. This photo step masks the p-well region (NMOS region). Resist is coated on SSI track using COATMTL.RCP recipe. Resist coated is thicker than the resist coated using COAT recipe.

The wafers are coated with HMDS prime before coating it with resist to improve adhesion. After coat, wafers are soft baked at 90ºC. The coated wafers are exposed on ASML stepper (exposure energy 250$mJ/cm^2$) using ADV NWELL mask.

The exposed wafers are developed in CD-26 developer and hard baked at 110ºC using DEVMTL.RCP recipe. After develop the overlay errors and resolution is inspected under a microscope and is measured to be 0 and 0.5$\mu m$ respectively.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

5.14 N-WELL IMPLANT

Ions are implanted in Varian 350D implanter. This is a medium current implanter; in this case N-type impurities are implanted. Phosphorus gas is used to extract $P_{31}$ carriers from the source.

Fig. 30: High energy NWELL implant.

The energy of the Monte Carlo implant should be high enough to go through 3500 Å thermal oxide to avoid surface inversion. The implant energy is 170 KeV and the implant dose is $5 \times 10^{13}$ cm$^{-2}$. Well dose, retrograde well dose and source/drain extension dose contributes to the threshold voltage of the device. Hence implant steps and drive in/anneal steps are critical part of MOSFET fabrication.

5.15 PHOTORESIST STRIP

Resist is stripped on GASONICS Asher, resist is stripped in O$_2$ plasma, the resist with implanted impurities is difficult to strip. It can be stripped using either FF recipe which has 2 heat steps one after every 98 seconds or 111 which has 3 heat steps after every 60 seconds. Recipe with three heat steps is ideal to strip the resist after high energy implants.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

5.16 THIRD LEVEL PHOTO (P WELL)

Third level lithography is P-well implant. This photo mask uses ADV P WELL mask which the opposite of NWELL mask. It mask’s the NWELL active area and P WELL active area is exposed after coat.

The coat recipe used is COATMTL.RCP; resist is thicker after coat. It is soft baked at 90ºC before exposure. Wafers are then exposed in ASML stepper with the energy 250 mJ/cm².

Wafers are then developed using CD-26 developer and they are inspected under microscope for overlay errors and resolution.

5.17 P-WELL IMPLANT

After photo, the wafers are implanted with p-type impurities. B11 is the impurity implanted for this process step. The source gas used is Boron. The implant energy is 110 KeV, since B11 impurity atom is smaller in size.

![Implant Profile](image)

Fig. 31: High energy P WELL implant.

Low energy B11 implant is enough to channel through the field oxide to avoid surface inversion. Boron dose implanted is $7 \times 10^{13} \text{ cm}^{-2}$. 
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.18 PHOTORESIST STRIP

Resist is stripped on GASONICS Asher in O$_2$ plasma, the resist with implanted impurities is difficult to strip. It can be stripped using either FF recipe which has 2 heat steps one after every 98 seconds or 111 which has 3 heat steps after every 60 seconds. Recipe with three heat steps is ideal to strip the resist after high energy implants.

5.19 WELL DRIVE-IN

Wafers undergo a thermal step after the implant to anneal implant damage. And drive in impurities to form P carriers dominated and N carriers dominated well, so that both PMOS and NMOS can be fabricated in the same wafer.

The furnace is warmed up using recipe 888 and the wafers are loaded in Tube 2 or Tube 3. Recipe 11 is then started which is the anneal recipe, this recipe has 27 minutes warm-up and stabilize step at 800°C. The temperature is then ramped up to 1000°C, and soaked at 1000°C for 4 hours in nitrogen ambient. It is followed by 40 minutes ramp down from 1000°C to 800°C and boat pull out.

P WELL

![Fig. 32: PWELL after drive-in.](image)
Simulation results show that the junction depth of PWELL is 1.4 µm and of NWELL is 1.08 µm. The diffusion co-efficient of boron and phosphorus is same in silicon but the mass of boron is less than phosphorus helping it to drive more into the silicon.

5.20 FOURTH LEVEL PHOTO (N-RETRO)

Fourth level lithography is for n-type retrograde well implant. The mask used for this implant is the same mask used for NWELL (ADV NWELL). And the recipe used for coat, expose and develop is similar to the one described in section 4.13.

5.21 N-RETROGRADE WELL IMPLANT

This implant step is critical in fabricating devices in this technology node. The purpose of retrograde well implant is to increase the concentration of impurities in the surface than in the bulk. Doing so will avoid short channel effects like punch-through and DIBL; it also minimizes the lateral diffusion of impurities to form a effective channel underneath the Polysilicon. The dose of the retrograde well implant is slightly higher than the well implant. This increases the concentration of impurities in the
Fabrication of 100 nm \( L_{\text{eff}} \) CMOS Devices

surface. The NTYPE (Phosphorus) dose implanted is \( 9.0 \times 10^{13} \text{ cm}^{-2} \). Impurities are implanted at 70KeV.

![Fig. 34: N-TYPE retrograde implant.](image)

The implant damage is annealed during gate oxide growth. The simulation profile of retrograde well is shown during Poly-Reoxidation.

### 5.22 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe) or recipe 111.

### 5.23 FIFTH LEVEL PHOTO (P-RETRO)

Fifth level lithography is for p-type retrograde implant. In this step, NWELL regions are masked with photoresist and PWELL regions are exposed. ADV PWELL mask is used for this level of lithography. The coat, expose and develop recipe used is similar to the one described in section 4.15.
Fabrication of 100 \( nm \) \( L_{eff} \) CMOS Devices

5.24 P-RETROGRADE WELL IMPLANT

B11 impurities are implanted again in PWELL with low energy, to increase the concentration of impurities in the surface of the silicon. Impurity dose implanted is \( 1.0 \times 10^{14} \, cm^{-2} \).

![Profile of B11 impurity concentration after retrograde well implant](image)

Fig. 35: P-TYPE retrograde implant.

B11 impurities are implanted with energy of 40\( KeV \). The profile of B11 impurity concentration after retrograde well implant is as seen in the figure above.

5.25 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe) or recipe 111.

5.26 PAD OXIDE ETCH

435 Å of screening oxide is etched prior to gate oxide growth. Etch rate of thermal oxide in 10:1 BOE is 560 \( \AA/min \) at room temperature. In this process, oxide in the active areas and oxide in field areas are etched.

After etch, wafers are rinsed in DI water for 5 minutes and is followed by SRD.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

5.27 RCA CLEAN

Prior to gate oxide growth, there are two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.

There are two short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 and SC2 to remove chemically grown oxide.

5.28 PRE-OXIDE ETCH TO GATE OXIDE GROWTH

Wafers go through a short HF dip again to ensure there is no native oxide prior to gate oxide growth.

5.29 GATE OXIDE GROWTH

A thin gate oxide recipe is developed for faster switching, high capacitance for high current drive. The only drawback of thin gate oxide is tunnelling of carriers from polysilicon into the channel.

The target thickness of the gate oxide growth for this process is 30 Å. Nitrogen is incorporated by introduction of nitrous oxide gas in the tube during gate oxide growth to prevent boron segregation into the gate oxide from P+ polysilicon or the P+ source/drain extensions during source drain anneal.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Gate oxide is grown in BRUCE Furnace Tube 4 dedicated for dry oxide growth. The tube is first warmed up to 800ºC using recipe 888. Then TransLC Clean recipe #463 is run to remove sodium impurities from the tube and ensure contamination free oxide is grown to avoid surface and trapped charges.

Fig. 36: Gate oxide growth.

After TransLC Clean, recipe 213 is loaded into the tube with three device wafers and two control wafers C1 and C2 in the boat. The recipe includes 20 minutes ramp up from 800ºC to 900ºC. It is followed by 10 minutes soak in dry O$_2$ and 10 minutes soak in N$_2$O at 900ºC. The furnace is ramped down and boat is pulled out.

The resulting thickness of the gate oxide is hard to measure on Nanospec or Spectrophotometer or on Ellipsometer. The dielectric behavior can be studied using CV measurements. For that capacitors have to be fabricated using different mask sets.

Oxynitride is used in SILVACO (deposited) simulation since thermal oxide cannot be grown using nitrous oxide as gas ambient.

Retrograde implants are anneal during this step and final retrograde profile is obtained during Polysilicon re-oxidation step.
**Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices**

### 5.30 LPCVD POLYSILICON DEPOSITION

Polysilicon is deposited on ASM LPCVD tool using low pressure chemical vapour deposition technique. It is deposited at 610°C, where the grain size is very small and reduces the channelling of impurities during source/drain implants. Polysilicon is grown using recipe FACPOLY610.

![Fig. 37: LPCVD polysilicon deposition.](image)

All three device wafers and two control wafers are loaded inside the tube. Silane ($\text{SiH}_4$) gas is used for deposition at a pressure of 300mTorr. The deposition time is calculated using the recent entries on log sheets. The deposition time is then calculated to be 37 minutes to deposit 2500Å of polysilicon at 610°C with a deposition rate of 67Å/min.

After the run the thickness of the polysilicon deposited is measured on spectrophotometer on a control wafer.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.31 SIXTH LEVEL PHOTO (POLY) – DOUBLE EXPOSURE

This photo step is one of the most critical steps to achieve the gate length of 150 nm. The minimum feature size on the mask is 0.25 µm. In this step, 1 µm device is double exposed to achieve a gate length of 150 nm.

The normal COAT.RCP and DEVELOP.RCP recipes are used for coat and develop respectively.

The mask used for this process step is JG POLY and the stepper job used is factory-double-exposure. The exposure energy used for both exposure steps are 190 mJ/cm$^2$. The direction of shift is defined in Y-axis and the units in cm. The job is defined such that wafer stage is shifted for a second pass.

After exposure, the overlay errors and resolution is inspected on the microscope.

5.32 POLYSILICON ETCH

Polysilicon is etched using RIE (Reactive Ion Etching) technique. The tool used for etching polysilicon is DryTek Quad. This tool is used to obtain anisotropic etch profiles with high aspect ratio (5:3 in this case).

![Fig. 38: Polysilicon etch.](image)
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Recipe details are described in Chapter 3 Section 3.4.

The control wafers are used to calculate the etch rate and SEM is used to see the cross section of the wafer and measure the gate length after etch.

**5.33 PHOTORESIST STRIP**

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

**5.34 RCA CLEAN**

Wafers undergo two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.

There is a short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 to remove chemically grown native oxide.

**5.35 POLYSILICON RE-OXIDATION**

Polysilicon re-oxidation is used as a stress relief layer for both silicon and polysilicon from following nitride deposition for spacers. It is also used as an etch stop layer during nitride spacer etch.

The device wafers and one control wafer go through a thermal oxide step where 250Å of dry oxide is grown using recipe 250.
The figure above shows the final retrograde profile of NMOS and PMOS respectively. The recipe in this thermal oxide growth step warms to 800ºC and followed by ramp up step from 800ºC to 1000ºC for 20 minutes. The wafers are then soaked in dry O₂ ambient for 45 minutes at 1000ºC. After ramp down and boat pull out, the oxide is measured on 5 different points on the wafer on nanospec and average thickness is measured to be 295Å.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

### 5.36 SEVENTH LEVEL PHOTO (P-SOURCE/DRAIN EXTENSION)

The seventh level lithography step is to mask NMOS regions with photoresist and open/expose PMOS regions for source drain extension implants. The mask set used for this job is JG PLDD and the job name selected on stepper is PLDD.

The normal COAT.RCP and DEVELOP.RCP recipes are used for coat and develop respectively. 1µm thick resist is coated with this recipe and it is thick enough to block low energy implants.

The energy used for exposure is 225$mJ/cm^2$. After develop the wafers are inspected underneath the microscope for resolution and overlay errors.

#### 5.37 (P+) – SOURCE/DRAIN EXTENSION IMPLANT

The source drain extensions and source drain regions are self aligned to gate. The implants are through thin screening oxide and implant energy is very low at 20KeV.

The impurity implanted is BF$_2$, because BF$_2$ is a large molecule gives shallow junctions (projected range and straggle of BF$_2$ is small compared to B11) and controlled lateral diffusion. BF$_2$ implant dose used in this process is $9.0\times10^{14} cm^{-2}$.

![PMOS, BF2 source drain extension implant](image)

Fig. 41: PMOS, BF$_2$ source drain extension implant.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

The implant dose is medium in order to decrease the width of depletion region extending into channel region and reducing source drain extension series resistance. Since Varian 350D is a medium current implanter, the implant time for implanting these device wafers with high dose and low energy is around 1 hour per wafer.

5.38 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

5.39 EIGHTH LEVEL PHOTO (N- SOURCE/DRAIN EXTENSION)

The eighth level lithography step is to mask PMOS regions with photoresist and open/expose NMOS regions for source drain extension implants. The mask set used for this job is JG NLDD and the job name selected on stepper is NLDD.

The normal COAT.RCP and DEVELOP.RCP recipes are used for coat and develop respectively. 1µm thick resist is coated using this recipe and it is thick enough to block low energy implants.

The energy used for exposure is $225mJ/cm^2$. After develop the wafers are inspected underneath the microscope for resolution and overlay errors.

5.40 (N+) - SOURCE/DRAIN EXTENSION IMPLANT

The impurity used for N-type source drain extension implant is Arsenic. Since arsenic is a large atom and has small projected range and straggle, Arsenic helps in achieving shallow junctions. The implant energy used for arsenic implants is also 20KeV and implant dose used is $5 \times 10^{15} cm^{-2}$. Arsenic implant dose is higher than BF$_2$ because arsenic is less effective to TED at lower temperature and requires higher energy
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

(temperature) than BF$_2$ to laterally diffuse under polysilicon. Higher dose compensates for temperature.

![Fig. 42: NMOS, Arsenic source drain extension implant.](image)

Higher dose also serves the purpose of smaller depletion widths and reduce parasitic resistance. Since RIT’s SMFL doesn’t have arsenic source for arsenic implants; the wafers should be sent out for source drain extension implants.

**5.41 PHOTORESIST SRTIP**

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

**5.42 RCA CLEAN**

Wafers undergo two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.

There is a short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 to remove chemically grown native oxide.

**5.43 LPCVD NIRIDE DEPOSITION**

Nitride is deposited using LPCVD technique. It is deposited on lower tube of ASM LPCVD. The temperature and gases to be used for deposition is initialized running a cassette in the tube.

Factory Nitride 810 recipe is used for deposition. The deposition temperature is 810°C, two gases Ammonia (NH$_3$-150 sccm) and Dichlorosilane (SiH$_2$Cl$_2$-60 sccm) is used for nitride deposition. Deposition time is calculated using the deposition rate from most recent nitride deposition run rate noted on the log sheet. In this case, nitride is deposited for 35 minutes with a base pressure of 300 mTorr.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

The thickness of the nitride layer deposited is measured using spectrophotometer using nitride on oxide recipe. The average thickness was measured to be $2749.7\text{Å}$.

The time, temperature and deposition rate is updated on the log sheet.

**5.44 NITRIDE ETCH**

After blanket deposition of nitride on the device wafers and one control wafer, nitride is etched using RIE technique on Drytek Quad tool.

![RIE nitride etch](image)

Sidewall spacers are formed on the sides of polysilicon after etching nitride. If the etch is completely anisotropic the width of the sidewall spacer is defined by thickness of the polysilicon. Nitride is first etched on control wafer to calculate the etch rate. Etch rate is calculated to be $1350\text{Å/.min}$. It is important to overetch, to ensure that no nitride is present on the source drain areas.

The gases used to etch nitride are $\text{SF}_6$ and $\text{CHF}_3$, the gas flows of each of the gases are $30\text{sccm}$ and the chamber pressure was maintained to be $40\text{mTorr}$. The RF power during etch is 250 Watts.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

The final spacer’s profile of the nitride and width of the nitride is measured in SEM.

### 5.45 NINTH LEVEL PHOTO (N-TYPE SOURCE/DRAIN)

Ninth level lithography is for N-type source drain implant. In this step, PMOS regions are masked with photoresist and NMOS regions are exposed. JG N+DS mask set is used for this level of lithography. The stepper job used is fac_adv_cmos and layer ID used is N+DS.

The coat, expose and develop recipe used is similar to the one described in section 4.38.

### 5.46 (N+) – SOURCE/DRAIN IMPLANT

Purpose of highly doped deeper junction source drain regions is for good ohmic contact. The gas source gas used is phosphorus and impurity is P31.

![NMOS (P31) source drain implant](image)

**Fig. 45:** NMOS (P31) source drain implant.

Implant energy and dose is 25KeV and $1.0 \times 10^{15} \text{ cm}^{-2}$ respectively. The reason why source drain dose is less than source drain extension dose for NMOS is the P31 is
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

used as an impurity. Diffusion of P31 in silicon is high compared to Arsenic. So, dose is compensated to make this device more capable of manufacturing in RIT and to reduce the cost of manufacturing. $1.0 \times 10^{15} \text{cm}^{-2}$ is still high dose and serves the purpose of making the source drain contact ohmic.

5.47 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

5.48 TENTH LEVEL PHOTO (P- SOURCE/DRAIN)

Tenth level lithography is for p-type source drain implant. In this step, NMOS regions are masked with photoresist and PMOS regions are exposed. JG P+DS mask set is used for this level of lithography. The stepper job used is fac_adv_cmos and layer ID used is P+DS.

The coat, expose and develop recipe used is similar to the one described in section 4.35.

5.49 (P+) - SOURCE/DRAIN IMPLANT

BF$_2$ is the impurity used for source drain implant. The source gas used is boron and implant dose is $5.0 \times 10^{15} \text{cm}^{-2}$ and the implant energy is 27KeV. The contact will be ohmic for the specified dose.

1 µm thick resist is thick enough to mask implants at this energy. BF$_2$ impurities are influenced to TED; hence spike annealing technique is used to anneal the implant damage.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Fig. 46: PMOS (BF2) source drain implant.

Implant time is 1 hour per wafer since implant is carried out at low energy on a medium current implanter.

5.50 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS Asher either using recipe FF (standard factory recipe).

5.51 RCA CLEAN

Wafers undergo two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

There is a short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 to remove chemically grown native oxide.

5.52 SOURCE/DRAIN SPIKE ANNEALING (RTA)

Silicon lattice is damaged due to ion implantation in previous steps. And dopants like BF$_2$ and Arsenic introduced into source drain regions needs to be electrically active and help in conduction.

More details regarding the preference of spike annealing over conventional furnace anneal is discussed in Chapter 3. To achieve shallow junctions and to electrically activate the dopants a thermal step is necessary, which is rapid thermal annealing in this case.

![Fig. 47: NMOS after spike anneal.](image)

AG610 RTP tool is used for annealing implant damages. This tool is capable of ramping up the temperatures from room temperature to 1050°C in 6 seconds avoiding lower temperatures where TED is significant. The soak time is 5 seconds in nitrogen ambient. And the tool ramps down from 1050°C to 600°C in 5 seconds.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

Fig. 48: Vertical cutline of NMOS source drain regions and their extensions with concentrations and junction depth.

The figure above shows the source drain junction’s profiles after anneal. The simulation results show that junction depth of source drain extensions is 40nm and junction depth of source drain regions is 102 nm.

Fig. 49: Horizontal cut line of NMOS in the channel region.

The figure above shows the horizontal cut line in the channel region giving an approximate value for effective channel length. The value obtained for NMOS $L_{eff}$ is 90 nm.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Fig. 50: PMOS after spike anneal.

The figure above shows PMOS after annealing. Most of the energy (temperature) during annealing is used in annealing the lattice damage hence dopant diffusion of implanted impurities is limited.

Fig. 51: Vertical cutline of PMOS source drain regions and their extensions with concentrations and junction depth.

The figure above shows the concentration of impurities in source drain extension regions and source drain regions. The junction depth of PMOS in source drain extension region and source drain regions is 50 nm and 85 nm respectively.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

Fig. 52: Horizontal cut line of PMOS in the channel region.

The Fig. 52 above shows the horizontal cut line in the channel region giving the value of effective channel length. The value obtained for PMOS is <80 nm.

5.53 OXIDE ETCH

The oxide over source drain regions is then etched in 50:1 HF for 2 minutes. It will ensure that there is no oxide in the active area and blanket titanium can be deposited to form Silicide.

5.54 TITANIUM DEPOSITION

Titanium is deposited using sputtering technique to form Silicide over source drain regions to reduce the sheet resistance. CVC601 is the sputter tool used to sputter titanium. The wafers are loaded inside the tool and the chamber is pumped down to a base pressure of $5 \times 10^{-5} \text{Torr}$. The sputter target is 4 inch and the power distributed on the target is 350 Watts. Argon is the sputtering gas used, the base pressure decides the mean free path length and the uniformity of titanium deposition.
The chamber goes through pre clean step before sputtering on wafers. After clean the shutters are opened and titanium is sputtered for 4 minutes to deposit a target thickness of 300 Å.

**5.55 RTA 1: TiSi (C49-PHASE)**

Titanium reacts with silicon and polysilicon to form Self Aligned Silicide. The high resistive titanium is formed at 650ºC with sheet resistance ranging between 5-7 Ω/□.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

The wafers go through the same AG610 RTP tool. The FACTiSi1 recipe has 5 seconds ramp up from room temperature to 650ºC. The wafers are soaked in nitrogen ambient for 5 seconds at 650ºC. And the temperature is ramped down from 650ºC to <100ºC in 5 seconds.

In this process the C54 phase TiSi is formed in the regions where titanium forms an alloy with silicon and polysilicon.

**5.56 TITANIUM ETCH (UNREACTED)**

Unreacted titanium is etched using hot plate chemistry. The solution is 1:2 H$_2$SO$_4$:H$_2$O$_2$ kept on hotplate at 150ºC. Each wafers are etched for 2 minutes to remove unreacted titanium. The Silicide does not etch in this chemistry, and the wafers are ready for second thermal anneal process.

**5.57 RTA 2: TiSi$_2$ (C54-PHASE)**

![Fig. 55: Titanium SALICIDE C54 phase.](image)

After unreacted titanium etch. The wafers go through a thermal process again in AG610 RTP tool. This time wafers are soaked at a higher temperature at 715ºC for 5 seconds to achieve much lesser resistivity of 4Ω/□. The least sheet resistance (3Ω/□)
Fabrication of 100 nm $L_{eff}$ CMOS Devices

is obtained at 850°C, but this temperature range is avoided where TED is enhanced and affects the source drain junctions.

**5.58 PECVD TEOS DEPOSITION**

![Fig. 56: PECVD TEOS deposition.](image)

In this step, TEOS is deposited as an inter-level dielectric. TEOS acts as an isolation layer from Metal 1 to polysilicon. 3000 Å of TEOS is deposited to give a uniform step coverage and would have helped to manufacture planar device if the CMP was available. It is also thick enough to handle the electric field breakdown.

TEOS is deposited on all three device wafers and on one control wafers to monitor the deposition rate on P5000 tool and to calculate the etch rate for contact cut etch step.

TEOS deposition is the choice over others for faster deposition rate at lower temperature. TEOS is deposited at 390°C for 61 seconds.
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.59 ELEVENTH LEVEL PHOTO (CONTACT CUT)

The eleventh level lithography step is to mask all regions with photoresist and open/expose the contact regions (part of source/drain region). The mask set used for this job is JG CC and the job name selected on stepper was CC.

The normal COAT.RCP and DEVELOP.RCP recipes are used for coat and develop respectively with resist thickness of 1µm.

The energy used for exposure is 225$mJ/cm^2$. After develop the wafers are inspected underneath the microscope for resolution and overlay errors.

5.60 CONTACT CUT ETCH

RIE technique is used to etch contact cuts. Contacts are etched on Drytek Quad using CH$_4$ (10sccm), CHF$_3$ (50sccm), O$_2$ (5sccm) and Argon (100sccm) gases. The RF power used to etch TEOS is 200 Watts and the pressure in the chamber is 100mTorr.

The etch time calculated to etch 4000Å of TEOS is 13 minutes with two minutes of over etch, TEOS was etched for 15 minutes. The selectivity of etch is very good in this case because the etch rate of underlying Silicide is 1Å/min.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

5.61 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

5.62 RCA CLEAN

Wafers undergo final clean step in two chemical baths; SC1 is the first bath and it contains the mixture of Ammonium Hydroxide (NH$_4$OH), Hydrogen Peroxide (H$_2$O$_2$) and DI water and the composition of this mixture is 1:1:16 respectively. Base mixture (SC1) helps in removing organic impurities.

SC2 is the second bath which contains the mixture of Hydrochloric Acid (HCl), Hydrogen Peroxide (H$_2$O$_2$) and DI water with 1:1:16 compositions. Acid mixture (SC2) helps in removing metal contaminants.

There is a short 30 seconds 50:1 (HF:H$_2$O) HF dip after SC1 to remove chemically grown native oxide.

5.63 ALUMINUM DEPOSITION

Aluminum is deposited using sputtering technique on CVC601 sputter tool. Aluminum target consists of 1% Si to avoid junction spiking and good adhesion with silicon. Aluminum target is 8 inch and it is deposited under a very low pressure of $1 \times 10^{-6}$ Torr. Deposition under such low pressure helps the uniformity across the wafer.

Such low pressure in the chamber is obtained with the help of rough pump and Cryo pump.
Fabrication of 100 nm $L_{eff}$ CMOS Devices

Fig. 58: Aluminum sputter deposition.

The power distributed to 8 inch target to sputter Aluminum is 2000 Watts. Argon is used as sputter gas and Argon flow is set to 17 sccm. The deposition time is 500 seconds to deposit 5000 Å of Aluminum with 1% Si.

5.64 TWELVETH LEVEL PHOTO (METAL 1)

Level twelve lithography is to define contact pads and hence areas of Aluminum etch are exposed to avoid shorts. The mask set used for this job is JG M1 and the job name selected on stepper was M1.

The normal COAT.RCP and DEVELOP.RCP recipes are used for coat and develop respectively with resist thickness of 1 µm.

The energy used for exposure is 225 mJ/cm². After develop the wafers are inspected underneath the microscope for resolution and overlay errors.

5.65 ALUMINUM ETCH

After the metal contacts are defined in the previous lithography step, Aluminum is plasma etched using chlorine ions. The tool used for this etch process is LAM4600.
Fabrication of 100 \textit{nm} \textit{L}_{\text{eff}} \text{CMOS} Devices

Plasma etch is desired to get anisotropic etch profiles. Wafers should rinsed in DI water after etching in Chlorine plasma to avoid corrosion. The chamber pressure maintained during the etch is 100\textit{mTorr}. The recipe details are shown in the Table 5.

Table 5: Aluminum etch recipe (LAM4600)

<table>
<thead>
<tr>
<th>STEP</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRESSURE</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>RF TOP (W)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RF BOTTOM</td>
<td>0</td>
<td>250</td>
<td>125</td>
<td>125</td>
<td>0</td>
</tr>
<tr>
<td>GAP (cm)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5.3</td>
</tr>
<tr>
<td>OXYGEN 111</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NITROGEN 222</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>BCI 333</td>
<td>50</td>
<td>50</td>
<td>25</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>CHLORINE 444</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>ARGON 555</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CFORM 666</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>COMPLETE</td>
<td>STABL</td>
<td>TIME</td>
<td>TIME</td>
<td>OVRETCH</td>
<td>TIME</td>
</tr>
<tr>
<td>TIME (sec)</td>
<td>15</td>
<td>8</td>
<td>150</td>
<td>10%</td>
<td>15</td>
</tr>
</tbody>
</table>
Fabrication of 100 nm $L_{\text{eff}}$ CMOS Devices

5.66 PHOTORESIST STRIP

Photoresist is stripped in oxygen plasma on GASONICS asher either using recipe FF (standard factory recipe).

5.67 SINTER

The wafers are sintered to ensure that the contact between Aluminum, Silicide and source drain regions is good. Good contact ensures low voltage drop in the linear region of operation of a transistor. Wafers are sintered at lower temperature ~450ºC and soaked in forming gas (H$_2$/N$_2$) for 15 minutes. Hydrogen reacts with the Si-SiO$_2$ interface and reduces oxide charges. These traps can increase parasitic resistance and hence increase threshold voltage of the device.
CHAPTER 6

ELECTRICAL CHARACTERISTICS OF SIMULATED DEVICE

In this chapter the simulation results and the reason for the failure of fabricated results is discussed.

6.1 SUPPLY VOLTAGE

As the thickness of gate oxide is scaled down with gate length, the supply voltage is also decreased to prevent oxide breakdown. Decrease in supply voltage also reduces leakage; the gate voltage is also scaled down. Hence, the transistor should be designed such that $V_T$ is much lower than the gate voltage. It is important because the drive current is proportional to the difference of gate voltage and threshold voltage ($V_{GS} - V_T$).

The contents discussed in this chapter are IV curves (Linear), DIBL curves; DIBL curves also give sub-threshold information, family of curves, $V_t$ roll-off and Terada Muta for both PMOS and NMOS. These results are obtained from ATLAS simulation.

Sheet resistance is compared between modeled and fabricated devices and failure of fabricated devices is discussed.
Simulated Electrical Characteristics

6.2 V-I curves (Linear)

NMOS

NMOS linear region of operation gives the threshold voltage of the device operated at drain bias at 0.1 V ($V_{DS} = 0.1 \, V$).

![Fig. 60: NMOS Linear IV curve.](image)

The gate is swept from 0V to 1.2V at drain bias 0.1V as mentioned above. The transistor is designed such that the device turns on well within the limit of gate voltage sweep. The NMOS I-V curve seen above is simulation result from ATHENA. Both recombination and mobility models are used; with gate contact defined as N+ poly for NMOS simulation. The interface charge concentration is assumed to be $3 \times 10^{11} \, cm^{-2}$. The threshold voltage of NMOS is modeled to be 0.4794V. This model is designed to reduce the effects of parasitic resistance on the performance of NMOS and PMOS.
Simulated Electrical Characteristics

PMOS

Even the PMOS linear region operation is obtained at drain bias ($V_{DS}$) equal to 0.1V.

![PMOS linear IV curve](image)

Fig. 61: PMOS linear IV curve.

The gate is swept from 0V to -1.2V, and the PMOS is designed such that threshold voltage is well within the gate voltage sweep. Like for NMOS both mobility and recombination model is used simulating IV curves. The interface charge concentration is assumed to be $3 \times 10^{11}$ cm$^{-2}$ and the contact defined is P+ Poly. The threshold voltage for this device is -0.58V.

The on current performance can be improved by reducing the dose of retrograde well implants, hence reducing the threshold voltage less than -0.58V.
Simulated Electrical Characteristics

6.3 DIBL (Drain Induced Barrier Lowering) CURVES

DIBL is important electrical characteristics to look at in sub micron regime. Since the drain bias plays an important role in lowering the barrier at source end and higher injection of carriers is observed in the channel region reducing the threshold voltage of the device at higher drain bias.

**NMOS**

![Logarithmic Curves](image)

The IV curves seen above are in logarithmic scale. The red curve is the IV curve at 0.1V drain bias and the green IV curve is at 1.2V drain bias. DIBL is the measure of the difference in threshold voltage at 0.1V drain bias and 1.2V drain bias.

DIBL is measured by sweeping gate from 0V to 1.2V at 0.1V drain bias and sweeping the gate again from 0.1V to 1.2V at 1.2V drain bias.

Fig. 62: NMOS logarithmic DIBL IV curve.
Simulated Electrical Characteristics

DIBL was measured to be $10.3\,mV/V$, and this program also includes recombination and mobility models. The current at $0\,V$ is $\sim 10^{-12} \, A/\mu m$. The designed device has a good OFF state performance compared to ON state performance.

PMOS

DIBL simulations for PMOS follow the same procedure as NMOS.

![Logarithmic Curves](image)

Fig. 63: PMOS logarithmic DIBL IV curve.

The gate is swept from $0\,V$ to $-1.2\,V$ at drain bias equal to $0.1\,V$; the DIBL curve in red is for the condition mentioned above ($V_{DS}=0.1\,V$). The DIBL curve in green follows the same sweep on gate with drain bias at $1.2\,V$.

The current at $0\,V$ is $\sim 10^{-14}\,A$; this modeled device has a good off state performance. The on state performance can be increased by reducing the implant dose for retrograde well in the channel region.
Simulated Electrical Characteristics

6.4 SUB-THRESHOLD SLOPE

Sub-threshold region of operation is a measure of current in off state. A transistor is in sub-threshold region of operation when gate bias is very much less than threshold voltage and surface/channel of the device is in weak inversion or depletion region, the corresponding current measured is called sub-threshold current.

Sub threshold slope is the measure of switching speed of the transistor from ON state to OFF state. Faster the device better is the performance.

6.5 FAMILY OF CURVES

NMOS

Fig. 64: NMOS family of curves.

The family of IV curves shown above for NMOS is a plot of drain current vs drain voltage swept from 0V to 1.2V at different gate voltages (0.3V to 1.2V). Linear region
Simulated Electrical Characteristics

of operation shows that the source and drain contacts are ohmic. The drive current measured at 1.2V drain bias on modeled NMOS device is 300µA/µm. Both recombination and mobility models are used to simulate the electrical characteristics.

The PMOS family of curves showed in Fig. 65, has a much lesser drive current than NMOS. The drive current at 1.2V for PMOS is 63µA/µm. The reason for degraded PMOS performance than NMOS is because mobility of electrons is greater than the mobility of holes in PMOS.

PMOS

![Fig. 65: PMOS family of curves.](image)

**6.6 VT ROLL-OFF**

Threshold voltage roll off is one the common short channel effect. Fig 66 is a plot of gate length versus threshold voltage. For gate lengths greater than 100 nm, threshold voltage is greater than designed value because more gate charge is required to turn the
Simulated Electrical Characteristics

transistor ON. As you can see in the figure at shorter gate lengths $V_t$ drops significantly.

This plot is obtained by simulating the NMOS and PMOS device at different gate lengths considering the design parameters used for 100 nm modeling.

![Vt Roll Off](image)

**Fig. 67: NMOS and PMOS $V_t$ Roll Off.**

### 6.7 TERADA MUTA ANALYSIS

Terada Muta is a method to extract the total sheet resistance and effective gate length of both PMOS and NMOS device. PMOS and NMOS are modeled for different gate lengths ranging from 60 nm to 500 nm. Resistance is measured on devices with different gate lengths by extracting current at $V_{GS} = 0.8V$, 1.0V, 1.2V and $V_{DS} = 0.1V$.

A plot of measured resistance vs gate length is obtained. Linear fit line is used to extrapolate the values from x-axis ($\Delta L$) and y-axis ($R_m$). Linear line fit does not give
Simulated Electrical Characteristics

accurate results for devices with very small gate lengths (<130 nm). So, the modeled device is further scaled down to 75 nm and 60 nm to obtain close to accurate results.

From the figure it can be said that $\Delta L$ is less than 60 nm for a device with mask length 150 nm. The $L_{eff}$ is therefore anywhere between 90 nm to 110 nm for NMOS. And the measured resistance is close to 500 to 600 $\Omega$.
Simulated Electrical Characteristics

The $\Delta L$ is less than 60 nm for a PMOS device with mask length 150 nm. And the measured resistance ranges between 2500 to 3500 $\Omega$. The effective gate length of the modelled device is less than 90 nm.

NOTE: Terada Muta analysis is done on shorter gate length since analysis on bigger gate lengths gave unacceptable results.

6.8 MODELED SHEET RESISTANCE

Table 6: Modeled Sheet Resistance.

<table>
<thead>
<tr>
<th>LAYER</th>
<th>SIMULATED SHEET RESISTANCE ($\Omega/\square$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P WELL</td>
<td>453</td>
</tr>
<tr>
<td>N+ SD Ext</td>
<td>296</td>
</tr>
<tr>
<td>N+ SD</td>
<td>93.7</td>
</tr>
<tr>
<td>N WELL</td>
<td>452.29</td>
</tr>
<tr>
<td>P+ SD Ext</td>
<td>2074.97</td>
</tr>
<tr>
<td>P+ SD</td>
<td>722.45</td>
</tr>
</tbody>
</table>

6.9 FAILURE OF FABRICATED DEVICES

Excessive gate leakage is noticed on the fabricated wafers. Gate current is measured in mAmps at 0.1V gate bias. The IV characteristics for PMOS looks like a resistor and for NMOS it looks like there is excessive gate leakage.

Fig. 70: IV curve of failed NMOS fabricated device.
As mentioned before, the gate leakage is the main reason for failed fabricated dies. Developing a good dielectric will help in successfully fabricating high performance modelled CMOS device.
CHAPTER 7

CONCLUSION AND FUTURE WORK

CONCLUSION

The entire process is designed to fabricate 100\textit{nm} or <100\textit{nm} L_{\text{effective}} devices. The software used to design this process is SILVACO. ATHENA and ATLAS simulators are the most used of the available simulators in SILVACO. Complex models are used in the design to match the fabricated results with results obtained on SILVACO. A device with good off state performance is modeled and simulated.

Unit process developments are made to the existing process to fabricate 100\textit{nm} CMOS devices. Parasitic resistance is reduced by heavily doping source drain extensions (LDD) and silicides are formed on source drain regions to form ohmic contact with very less forward voltage drop. Retrograde well implants are included in the process to avoid short channel effects. Spike annealing is used in place of conventional furnace source drain annealing to minimize the contribution of TED to enhanced source drain diffusion and limiting the thermal budget. Techniques like double exposure is developed to obtain the mask length feature size of 150\textit{nm}.

Device fabricated failed due to excessive tunnelling of charges from gate to the channel region. The 30Å thick gate oxide (thermal oxide with nitrogen incorporation) grown during the process seemed like there was no gate oxide grown due electrical testing. It is time to use high k dielectric gate oxide (hafnium oxide) to stop tunnelling of charges from gate to channel region.

Modeled on state performance of the device can be improved by reducing the dose of retrograde well and well implants.
Conclusion and Future Work

Modeled devices have very less drive current ~300 $\mu$Amps/$\mu$m for NMOS and ~63 $\mu$Amps/$\mu$m for PMOS; reflecting good off state performance.

This process is updated on MES system for students in microelectronic manufacturing to improve (develop a process to deposit thin high k gate dielectric) and run the entire process and electrically test the device successfully.

This research is funded by Dr. Lynn Fuller and Microelectronics department of RIT.

**FUTURE WORK**

This process can be improved by developing new recipe for gate oxide growth; use of high k dielectric is desirable. Stained silicon can be used for improved on state performance of the device. Use of flash anneal over spike anneal will give more control over dopant diffusion and the device can be further scaled to smaller size.

Titanium Silicide should be replaced by Nickel Silicide or metal with low barrier height. Advantage of chrome over Titanium is; Silicide is formed at much lower temperature ~300ºC and forward voltage drop is lower. Use of chrome gives more linear curves, in the linear region of operation.
APPENDIX

ATHENA NMOS PROGRAM

STRUCTURE 1

# Modelling and Simulation of 100nm effective channel length NMOS transistor

# Grid parameter definition for both x and y co-ordinates
line x loc=0.00 spac=0.1
line x loc=0.8 spac=0.01
line x loc=1.0 spac=0.01
#
line y loc=0.00 spac=0.01
line y loc=0.1 spac=0.01
line y loc=0.2 spac=0.1
line y loc=0.5 spac=0.1

# Diffusion model used is Fully coupled model
method full.cpl
# Initialising the mesh with phosphorous impurities making the starting wafer P-type
init silicon boron resistivity=20 orientation=100 two.d

# Grow pad oxide in dryo2 ambient
deposit oxide thick=0.043

# pwell implant with bf2 as an impurity
implant boron dose=7.0e13 energy=100 monte n.ion=100000 rotation=45 \crystal unit.damage dam.factor=0.05

# well drive in
diffus time=20 temp=800 t.final=1000 nitro
diffus time=240 temp=1000 nitro
diffus time=20 temp=1000 t.final=800 nitro

# extract the junction depth of pwell
extract name="junction_depth_of_pwell" xj material="Silicon" mat.occn=1 x.val=0.35 \junc.occn=1

# implant boron dose=1.0e14 energy=45 monte rotation=45 crystal
extract name="p_well_rs" sheet.res material="Silicon" mat.occn=1 x.val=1.00 \region.occn=1

tonyplot
struct outfile=100nmos_rev1.str
quit

STRUCTURE 2

# load the previous structure file
init infile=100nmos_rev1.str
Appendix

# etch pad oxide
etch oxide dry thick=.0485
#tonyplot

# gate oxide growth
deposit oxynitride thick=0.003
#
diffus time=10 temp=800 t.final=900 nitro
diffus time=20 temp=900 nitro
diffus time=10 temp=900 t.final=800 nitro
#tonyplot

# deposit polysilicon
deposit polysilicon thick=0.25 dy=0.02

# etch polysilicon start x=0 y=0.4
etch cont x=0 y=-0.6
etch cont x=0.9375 y=-0.6
etch done x=0.9375 y=0.4

# etch oxynitride left p1.x=0.9375

# poly reox
deposit oxide thick=0.025
#diffus time=20 temp=800 t.final=1000 nitro
diffus time=15 temp=1000 nitro
#diffus time=20 temp=1000 t.final=800 nitro
#tonyplot
#
struct outfile=100nmNMOS2.str
quit

STRUCTURE 3

go athena
method full.cpl
# load the previous structure file
init infile=100nmNMOS2.str

# ldd arsenic implant
implant arsenic dose=5.0e15 energy=20 monte n.ion=100000 smooth=0.25 \ rotation=45

# nitride spacers
deposit nitride thick=0.25

# etch nitride to form sidewall spacers
etch nitride dry thick=0.25

# source drain implants
implant phosphor dose=1.0e15 energy=25 monte n.ion=100000 smooth=0.25 \ rotation=45

#anneal
diffus time=3/60 temp=700 t.final=1050 nitro
diffus time=5/60 temp=1050 nitro
diffus time=6/60 temp=1050 t.final=700 nitro
#struct mirror right
#tonyplot
etch oxide dry thick=0.025
struct outfile=100nmNMOS3.str
Appendix

Tonyplot
quit

**STRUCTURE 4**

```plaintext
go athena
#
init infile=100nmNMOS3.str
#
deposit titanium thick=0.03 divisions=10 dy=0.015
#
diffuse time=5/60 temp=700 nitro
#
etch titanium all
#
diffuse time=5/60 temp=700 nitro
#
deposit oxide thick=0.30
#
etch oxide start x=.20 y=1.00
etch cont x=.20 y=-1.00
etch cont x=.40 y=-1.00
etch done x=.40 y=1.00
#
etch oxide start x=.95 y=1.00
etch cont x=.95 y=-1.00
etch cont x=1.0 y=-1.00
etch done x=1.0 y=1.00
#
etch oxide above p1.y=-0.275
struct mirror right
struct outfile=100nmNMOS4.str
tonyplot
quit
```

**STRUCTURE 5**

```plaintext
go athena
init infile=100nmNMOS4.str
#
deposit aluminum thick=0.40 divisions=10
#
#tonyplot
#
etch aluminum above p1.y=-0.26
tonyplot
#quit

electrode name=gate x=1.0 y=0.1
electrode name=source x=0.3
electrode name=drain x=1.7
electrode name=substrate backside
struct outfile=nmosvt.str
quit
```

**EXTRACTS**
Appendix

NMOS LINEAR IV CURVES

go atlas simflags="-P 1"
init infile=nmosvt.str

#set material models
models cvt srh print
contact name=gate n.poly
interface qf=3e11

#method newton
solve init

#Bias the drain
solve vdrain=0.1

#Ramp the gate
log outf=mos1ex01_1.log master
solve vgate=0 vstep=0.01 vfinal=1.2 name=gate
save outf=mos1ex01_1.str

#plot results
tonyplot mos1ex01_1.log -set mos1ex01_1_log.set
#extract device parameters
extract name="nvt" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain"))))
   - abs(ave(v."drain"))/2.0)
quit

NMOS DIBL

go atlas simflags="-P 1"
#
init infile=nmosvt.str

#OUTPUT CON.BAND

# set material models
models cvt srh print
contact name=gate n.poly
interface qf=3e11

# get initial solution
solve init

# Bias the drain a bit...
solve vdrain=0.1

# Ramp the gate
log outf=wxyz.log
solve vgate=0 vstep=0.01 vfinal=1.2 name=gate
#tonyplot wxyz.log

# extract device parameters
extract init inf="mos1ex04_1.log"
#extract names="nvt" x.val from curve(abs(v."gate"),abs(i."drain"))
extract name="nvt1" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain"))))
   - abs(ave(v."drain"))/2.0)

# now open a dummy log file...
log off

# Now start again and ramp the drain to 1.5 volts...
solve init
# Bias the drain to 1.2 volts......slowly at first....
Appendix

solve vdrain=0.025 vstep=0.01 vfinal=0.1 name=drain
solve vdrain=0.25 vstep=0.01 vfinal=1.2 name=drain
#tonyplot

# Ramp the gate again with another opened logfile...
log outf=abc.log
solve vgate=0 vstep=0.025 vfinal=1.2 name=gate

# extract the next device parameter with the drain now at 1.2 volts....
extract init inf="abc.log"
extract name="nvt2" x.val from curve(abs(v."gate"),abs(i."drain"))
extract name="nvt2" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) \ - abs(ave(v."drain"))/2.0)

# Calculate a DIBL parameter...in V/V
extract name="ndibl" ($"nvt1"-"nvt2")/(1.2-0.1)
tonyplot -overlay wxyz.log abc.log -set mos1ex04_log.set
quit

NMOS FAMILY OF CURVES

go atlas2 simflags="-P 1
init inf=pmosvt.str

# define the Gate workfunction
contact name=gate n.poly

# Define the Gate Qss
interface qf=3e11

# Use the cvt mobility model for MOS
models cvt srh

# set gate biases with Vds=0.0
solve init
solve vgate=0.3 outf=solve_tmp1
solve vgate=0.4797 outf=solve_tmp2
solve vgate=0.9 outf=solve_tmp3
solve vgate=1.2 outf=solve_tmp4

#load in temporary files and ramp Vds
load infile=solve_tmp1
log outf=mos1ex09_1.log
solve name=drain vdrain=0 vfinal=1.2 vstep=0.1

load infile=solve_tmp2
log outf=mos1ex09_2.log
solve name=drain vdrain=0 vfinal=1.2 vstep=0.1

load infile=solve_tmp3
log outf=mos1ex09_3.log
solve name=drain vdrain=0 vfinal=1.2 vstep=0.1

load infile=solve_tmp4
log outf=mos1ex09_4.log
solve name=drain vdrain=0 vfinal=1.2 vstep=0.1

# extract max current and saturation slope
extract name="pidsmax" max(abs(i."drain"))
extract name="p_sat_slope" slope(minslope(curve(abs(v."drain"),abs(i."drain"))))
tonyplot -overlay mos1ex09_1.log mos1ex09_2.log mos1ex09_3.log mos1ex09_4.log -set mos1ex09_1.set
quit
Appendix

ATHENA PMOS PROGRAM

STRUCTURE 1

go athena

# Modelling and Simulation of 100nm effective channel length PMOS transistor

# Grid parameter definition for both x and y co-ordinates
line x loc=0.00 spac=0.1
line x loc=0.8 spac=0.01
line x loc=1.0 spac=0.01
#
line y loc=0.00 spac=0.001
line y loc=0.1 spac=0.01
line y loc=0.2 spac=0.1
line y loc=0.5 spac=0.1

# Diffusion model used is Fully coupled model
method full.cpl

# Initialising the mesh with phosphorous impurities making the starting wafer n-type
init silicon boron resistivity=20 orientation=100 two.d
#
deposit oxide thick=0.043
#tonyplot

# nwell implant with phos as an impurity
implant phosphor dose=2.0e13 energy=170 monte rotation=45 crystal

# well drive in
diffus time=20 temp=800 t.final=1000 nitro
diffus time=240 temp=1000 nitro
diffus time=20 temp=1000 t.final=800 nitro

# extract the junction depth of pwell
extract name="junction_depth_of_pwell" xj material="Silicon" mat.occno=1 x.val=0.35 \ junc.occno=1
#
implant phosphor dose=7.0e13 energy=68 monte rotation=45 crystal

extract name="boron_vt_adjust" area from curve(depth,impurity="Net Doping" \ material="Silicon" mat.occno=1 y.val=0.2)/10000 outfile="vtadjust.dat"

# struct mirror right
#tonyplot
#quit
struct outfile=100pmos_rev1.str
quit

STRUCTURE 2

go athena
method full.cpl

# load the previous structure file
init infile=100pmos_rev1.str

# etch pad oxide
etch oxide dry thick=0.0485
#tonyplot
Appendix

# gate oxide growth
deposit oxynitride thick=0.003 divisions=5
diffus time=10 temp=800 t.final=900 nitro
diffus time=20 temp=900 nitro
diffus time=10 temp=900 t.final=800 nitro

# deposit polysilicon
deposit polysilicon thick=0.25 dy=0.02
#tonyplot
#
etch polysilicon start x=0 y=0.4
etch cont x=0 y=-0.6
etch cont x=0.9375 y=-0.6
etch done x=0.9375 y=0.4
#tonyplot
#
etch oxynitride left p1.x=0.9375
#
deposit oxide thick=0.025
#tonyplot
diffus time=20 temp=800 t.final=1000 nitro
diffus time=15 temp=1000 nitro
diffus time=20 temp=1000 t.final=800 nitro
#tonyplot
struct outfile=100nmPMOS2.str
quit

STRUCTURE 3

go athena
method full.cpl

# load the previous structure file
init infile=100nmPMOS2.str

# ldd implant
implant bf2 dose=9.0e14 energy=20 monte n.ion=100000 smooth=0.25 \rotation=45

# nitride spacers
deposit nitride thick=0.25

# etch nitride to form sidewall spacers
etch nitride dry thick=0.25
#tonyplot
# source drain implants
.implant bf2 dose=5.0e15 energy=28 monte n.ion=100000 smooth=0.25 \rotation=45

#anneal
diffus time=3/60 temp=700 t.final=1050 nitro
#
diffus time=5/60 temp=1050
#
diffus time=6/60 temp=1050 t.final=700 nitro
struct mirror right
#tonyplot
quit

# etch oxide
etch oxide dry thick=0.025
struct outfile=100nmPMOS3.str
#tonyplot
quit
Appendix

**STRUCTURE 4**
go athena
init infile=100nmPMOS3.str

#
deposit titanium thick=0.03 divisions=10 dy=0.015
#
diffuse time=5/60 temp=700 nitro
#
etch titanium all
#
diffuse time=5/60 temp=700 nitro
#
deposit oxide thick=0.30
struct mirror right
tonyplot

#
etch oxide start x=.20 y=1.00
etch cont x=.20 y=-1.00
etch cont x=.40 y=-1.00
etch done x=.40 y=1.00
#
etch oxide start x=.95 y=1.00
etch cont x=.95 y=-1.00
etch cont x=1.05 y=-1.00
etch done x=1.05 y=1.00
#
etch oxide start x=1.6 y=1.00
etch cont x=1.6 y=-1.00
etch cont x=1.8 y=-1.00
etch done x=1.8 y=1.00

etch oxide above p1.y=-0.28
struct outfile=100nmPMOS4.str
tonyplot
quit

**STRUCTURE 5**
go athena
init infile=100nmPMOS4.str

#
deposit aluminum thick=0.4 divisions=10
#
#tonyplot
#
etch aluminum above p1.y=-0.265
#tonyplot

electrode name=gate x=1.0 y=0.1
electrode name=source x=0.3
electrode name=drain x=1.7
electrode name=substrate backside
#tonyplot
struct outfile=pmosvt.str
quit
Appendix

**PMOS LINEAR IV CURVES**

go atlas simflags="-P 1"
init infile=pmosvt.str
#tonyplot
100nmPMOS3.str

#set material models
models cvt srh print
contact name=gate p.poly
interface qf=3e11

# get initial solution
solve init
method newton trap
solve prev

# Bias the drain a bit...
solve vdrain=-0.1 name=drain

# Ramp the gate
log outf=mos1ex08_1.log master
solve vgate=0 vstep=-0.01 vfinal=-1.2 vdrain=-0.1 name=gate
save outf=mos1ex08_1.str

# extract device parameters......
extract init inf="mos1ex08_1.log"
extract name="pvt" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) - abs(ave(v."drain"))/2.0)

# plot results
tonyplot   mos1ex08_1.log -set mos1ex08_1_log.set

**PMOS DIBL**

go atlas simflags="-P 1"
init infile=pmosvt.str
# set material models
models cvt srh print
contact name=gate p.poly
interface qf=3e11

# get initial solution
solve init
method newton trap maxtraps=8 autonr itlimit=30
solve prev

# Bias the drain a bit...
solve vdrain=-0.025 vstep=-0.025 vfinal=-0.1 name=drain

# Ramp the gate
log outf=mos1ex11_1.log master
solve vgate=0 vstep=-0.01 vdrains=-0.1 vfinal=-1.2 \ 
   name=gate cname=drain

# extract device parameters
extract init inf="mos1ex11_1.log"
extract name="pvt1" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) \ 
   - abs(ave(v."drain"))/2.0)

# now open a dummy log file...
log off

# Now start again and ramp the drain to -1.2 volts...
solve init
# Bias the drain to 3 volts......slowly at first....
solve vdrain=-0.025 vstep=-0.025 vfinal=-0.1 name=drain
solve vdrain=-0.25 vstep=-0.025 vfinal=-1.2 name=drain

# Ramp the gate again with another opened logfile....
Appendix

log outf=mos1ex11_2.log master
solve vgate=0 vstep=-0.01 vdrain=-1.2 vfinal=-1.2 name=gate \
cname=drain

# extract the next device parameter with the drain now at -1.2 volts....
extract init inf="mos1ex11_2.log"
extract name="pvt2" x.val from curve(abs(v."gate"),abs(i."drain")) where y.val=0.1e-6
eextract name="pvt2" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) \ 
- abs(ave(v."drain"))/2.0)

# Calculate a DIBL parameter...in V/V
extract name="pdiff" ($"pvt1"-"pvt2)/(1.2-0.1)
tonyplot -overlay mos1ex11_1.log mos1ex11_2.log -set mos1ex11_log.set
quit

PMOS FAMILY OF CURVES

go atlas2 simflags="-P 1"

init infle=pmosvt.str

# define the Gate workfunction
contact name=gate p.poly

# Define the Gate Qss
interface qf=3e11

# Use the cvt mobility model for MOS
models cvt srh

# set gate biases with Vds=0.0
solve init
solve vgate=-0.3 outf=solve_tmp1
solve vgate=-0.6 outf=solve_tmp2
solve vgate=-0.9 outf=solve_tmp3
solve vgate=-1.2 outf=solve_tmp4

# load in temporary files and ramp Vds
load infile=solve_tmp1
log outf=mos1ex09_1.log
solve name=drain vdrain=0 vfinal=-1.2 vstep=-0.1

load infile=solve_tmp2
log outf=mos1ex09_2.log
solve name=drain vdrain=0 vfinal=-1.2 vstep=-0.1

load infile=solve_tmp3
log outf=mos1ex09_3.log
solve name=drain vdrain=0 vfinal=-1.2 vstep=-0.1

load infile=solve_tmp4
log outf=mos1ex09_4.log
solve name=drain vdrain=0 vfinal=-1.2 vstep=-0.1

# extract max current and saturation slope
extract name="pidsmax" max(abs(i."drain"))
extract name="p sat_slope" slope(minslope(curve(abs(v."drain"),abs(i."drain"))))
tonyplot -overlay mos1ex09_1.log mos1ex09_2.log mos1ex09_3.log mos1ex09_4.log -set mos1ex09_1.set
quit
Reference

REFERENCE


[3] NTRS-Roadmap

[4] Stanford


